

Schweregrad- differenzierung in der neurologischen und Trauma-rehabilitation

Messinstrumente bei Schlag-
anfall und Schädel-Hirn-
Trauma

Endbericht Teil 1



Ludwig Boltzmann Institut
Health Technology Assessment

LBI-HTA Projektbericht Nr.: 023a
ISSN: 1992-0488
ISSN-online: 1992-0496

Schweregrad- differenzierung in der neurologischen und Trauma-rehabilitation

Messinstrumente bei Schlag-
anfall und Schädel-Hirn-
Trauma

Endbericht Teil 1



Ludwig Boltzmann Institut
Health Technology Assessment

Wien, Juli 2009

Projektleitung:	Priv.Doz.Dr.phil. Claudia Wild
Projektbearbeitung:	Christoph Pammer, DSA, MPH
Projektmitarbeit:	Muna Abuzahra, MSC
Systematische Literatursuche:	Tarquin Mittermayr, BA (Hons)
Externer Review:	Dr. med. Peter Wipfler

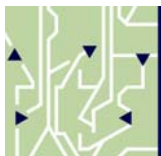
Dieser Bericht soll folgendermaßen zitiert werden/This report should be referenced as follows:

Pammer, C., Abuzahra, M., Wild, C. Schweregraddifferenzierung in der Neuro- und Traumarehabilitation. Messinstrumente bei Schlaganfall und Schädel-Hirn-Trauma. Systematischer Review. HTA Projektbericht 2009; 23a

IMPRESSUM

Medieninhaber und Herausgeber:
Ludwig Boltzmann Gesellschaft GmbH
Operngasse 6/5. Stock, A-1010 Wien
<http://www.lbg.ac.at/de/lbg/impressum>

Für den Inhalt verantwortlich:



Ludwig Boltzmann Institut für Health Technology Assessment (LBI-HTA)
Garnisongasse 7/20, A-1090 Wien
<http://hta.lbg.ac.at/>

Die LBI-HTA-Projektberichte erscheinen unregelmäßig und dienen der Veröffentlichung der Forschungsergebnisse des Ludwig Boltzmann Instituts für Health Technology Assessment.

Die Berichte erscheinen in geringer Auflage im Druck und werden über das Internetportal „<http://eprints.hta.lbg.ac.at/>“ der Öffentlichkeit zur Verfügung gestellt:

LBI-HTA Projektbericht Nr.: 023a
ISSN: 1992-0488
ISSN-online: 1992-0496

© 2009 LBI-HTA – Alle Rechte vorbehalten

Inhaltsverzeichnis

Inhaltsverzeichnis	3
Zusammenfassung	5
Summary	7
1 Einleitung	9
1.1 Ziel und Wissenschaftliche Fragestellungen	9
2 Methodik zur Beurteilung der Messinstrumente	13
2.1 Messtheorie	13
2.1.1 Validität	13
2.1.2 Reliabilität	14
2.1.3 Änderungssensitivität	14
2.2 Operationalisierung der Evaluationskriterien	16
3 Methodik: Systematische Literaturrecherche	19
3.1 PICO-Frage und Suchstrategie	19
3.2 Literaturauswahl	20
4 Schlaganfall	21
4.1 Schweregraddifferenzierung	21
4.1.1 National Institute of Health Stroke Scale (NIHSS)	22
4.1.2 Die Orpington Prognostic Scale (OPS)	23
4.1.3 Die Glasgow Coma Scale (GCS)	24
4.1.4 Modified Ranking Scale (mRS)	24
4.1.5 Kombinierte Testbatterien	25
4.2 Outcome measures/Ergebnismessung	26
4.2.1 Beck Depression Inventory (BDI)	26
4.2.2 Stroke Impact Scale Version 3.0 (SIS 3.0)	27
4.2.3 Barthel Index (BI)	28
4.2.4 Functional Independence Measure (FIM)	30
4.2.5 Frenchay Activities Index (FAI)	33
4.2.6 Mini-Mental State Examination (MMSE)	33
4.2.7 Berg Balance Scale (BBS)	34
4.2.8 Fugl-Meyer Assessment of Motor Recovery after Stroke (FMA)	35
4.2.9 Medical Outcomes Study Short Form 36 (SF-36)	36
4.2.10 Stroke Specific Quality of Life Scale (SSQOL)	37
4.2.11 Die EuroQol Quality of Life Scale (EQ-5D)	37
4.2.12 Nottingham Health Profile (NHP)	38
4.3 Zusammenfassende Beurteilung von Skalen und Scores zur Schweregraddifferenzierung und Outcome Measurement bei Schlaganfall	42
5 Schädel-Hirn-Traumata (SHT)	45
5.1 Schweregraddifferenzierung	45
5.1.1 Glasgow Coma Scale (GCS)	45
5.1.2 Revised Trauma Score (RTS)	45
5.1.3 Trauma and Injury Severity Score (TRISS _{RTS})	45
5.1.4 Schweregraddifferenzierung bei Kindern und Jugendlichen	46
Pediatric Trauma Score (PTS)	47
Pediatric Risk Indicator (PRI)	47
5.1.5 Prognosepotential von Instrumenten der Schweregraddifferenzierung	47
5.2 Outcome Scales	48
5.2.1 Glasgow Outcome Scale (GOS)	48

5.2.2	Rivermead Post-Concussion Symptoms Questionnaire (RPQ)	49
5.2.3	Barthel Index (BI)	50
5.2.4	Functional Independence Measure (FIM)	50
5.2.5	Functional Assessment Measure (FAM)	51
5.2.6	Disability Rating Scale (DRS)	52
5.2.7	Functional Status Examination (FSE)	53
5.2.8	Independent Living Scale (ILS)	53
5.2.9	Community Integration Questionnaire (CIQ).....	54
5.3	Patient/innen-Zufriedenheit	55
5.4	Exkurs: Rückenmarksverletzungen	55
5.5	Zusammenfassende Beurteilung von Skalen und Scores zur Schweregraddifferenzierung und Outcome measurement bei Schädel-Hirn-Traumata	58
6	Ausblick	61
	Quellen	63

Abbildungsverzeichnis

Abb. 3.2-1:	Darstellung des Auswahlprozesses (QUORUM tree)	20
Abb. 4.1-1:	Darstellung der Ergebnisse von der Schweregraddifferenzierung zur Ergebnismessung: entlang der ICF-Dimensionen	21
Abb. 4.2-1:	Vergleich der Testgüte verschiedener Depressionsskalen	27
Abb. 4.2-2:	Vergleich unerwünschter Informationseffekte von Barthel Index und Functional Independence Measure	32
Abb. 4.2-3:	Unerwünschter Informationseffekte bei verschiedenen Balance-Messungen	36
Abb. 4.3-1:	Zusammenfassende Beurteilung der Skalen und Scores bei Schlaganfall.....	42
Abb. 5.1-1:	Testgüte verschiedener Instrumente der Schweregraddifferenzierung bei Schädel-Hirn-Traumata	46
Abb. 5.1-2:	Testgütevergleich der Disability Rating Scale (Anm.: hier als Referenz für den „wahren prognostischen Wert“ mit anderen Skalen der Schweregraddifferenzierung)	48
Abb. 5.5-1:	Zusammenfassende Beurteilung der Skalen und Scores bei Schädel-Hirn-Trauma	58

Tabellenverzeichnis

Tab. 2.2-1:	Übersicht zur Operationalisierung der Evaluationskriterien	18
Tab. 3.1-1:	PICO-Suchstrategie	19
Tab. 4.2-4:	Übersicht der Skalen und Scores bei Schlaganfall	40
Tab. 5.4-1:	Übersicht der Skalen und Scores bei Schädel-Hirn-Traumata	56

Zusammenfassung

Die Evidenz, dass die Implementierung von struktur- und prozessgesteuerten Systemen zur Qualitätssicherung der neurologischen Rehabilitation eine Verbesserung der funktionalen Gesundheit der Patient/innen mit sich bringt ist begrenzt. Diese Erkenntnis legt nahe, auch zum Zwecke des Qualitätsmanagements nach Indikatoren zu suchen, die eine ergebnisorientierte Messung der funktionalen Gesundheit ermöglichen. Da die neurologische Rehabilitation darauf abzielt, nicht nur körperliche Funktionen und Organfunktionen wiederherzustellen, sondern den Patienten/innen auch die gesellschaftliche Teilhabe wieder zu ermöglichen, wurden neben einfachen Funktions-Scores auch eine Vielzahl von Instrumenten zur Beurteilung der Aktivität und der Partizipation entwickelt und validiert. Eine umfangreiche, mehrdimensionale Beurteilung des Gesundheitszustands bildet eine ausreichende Grundlage, um optimierende Allokationsentscheidungen in der Rehabilitation zu treffen, eine standardisierte Rehabilitationsplanung zu ermöglichen und um den Erwartungen der Patient/innen und Kostenträger an das Ergebnis der Rehabilitation gerecht zu werden.

Im ersten Teil des Projektes zur Schweregradifferenzierung in der Neuro- und Traumarehabilitation sollen Messinstrumente zur Schweregrad- und Ergebnisbewertung in der Neuro- und Traumarehabilitation anhand von zwei ausgewählten Diagnosegruppen (Schlaganfall und Schädel-Hirn-Trauma) dargestellt sowie deren Testgüte bewertet werden.

Mittels systematischer Literatursuche und Handsuche wurden 2527 Publikationen identifiziert, von denen auf Basis der Einschlusskriterien 167 Volltext-Artikel in die Analyse eingeschlossen wurden.

Zur Schweregraddifferenzierung und Ergebnisbeurteilung bei Schlaganfall weisen diagnosespezifische Messinstrumente bessere Testgüte als generische (diagnoseübergreifende) Instrumente auf. Aufgrund der Beurteilung nach den Testgütekriterien können die National Institut of Health Stroke Scale (NIHSS), der Beck Depression Inventory (BDI), der Frenchay Activities Index (FAI), die Stroke Impact Scale (SIS) und die Stroke Specific Quality of Life Scale (SSQOL) empfohlen werden. Bei den Instrumenten zur Schweregraddifferenzierung sowie Outcomeevaluierung bei Schädel Hirn Traumata zeigt sich, dass diagnosespezifische Messinstrumente nicht notwendigerweise die bessere Testgüte aufweisen. Hinsichtlich der Testgüte sind die Disability Rating Scale (DRS) und der Community Integration Questionnaire (CIQ) zu empfehlen.

Die Entscheidung für ein Messinstrument ist jedoch nicht alleine von deren Testgüte (und der Machbarkeit und Akzeptanz der Messinstrumente) abhängig, sondern vom Zweck, den Messungen in Mess-Systemen dienen sollen. Ein generisches Kerninstrument könnte diagnoseübergreifend zur Anwendung kommen, während andere Parameter zusätzlich diagnosespezifisch erfasst werden müssen (modulare Zusammensetzung von generischen und diagnosespezifischen Instrumenten).

Zur Auswahl von Messsystemen zur Unterstützung der neurologischen Rehabilitation in Österreich, müssen daher zunächst die Zielsetzungen der Messungen klar definiert werden und festgelegt werden, ob und in welcher Form Messungen zur Qualitätssicherung und Ergebnismessung eingesetzt werden sollen. Weiters sollte überlegt werden, ob und in welcher Form Mes-

Umfassende Beurteilung des Gesundheitszustandes von PatientInnen mit neurologischer Rehabilitation notwendig für Allokationsentscheidungen, Rehabilitationsplanung und PatientInnenzufriedenheit

Identifikation und Bewertung von Messinstrumenten

167 Volltext Artikel eingeschlossen

Schlaganfall: diagnosespezifische Instrumente weisen bessere Testgüte auf

Schädel-Hirn Trauma: generische Instrumente vorzuziehen

Modulare Zusammensetzung generischer und diagnosespezifischer Instrumente ev. notwendig

Vor Auswahl der Instrumente muss Zieldefinition der Messung erfolgen

sungen mit der Refundierung und den entsprechenden ökonomischen Anreizsystemen kombiniert werden sollen.

Summary

The evidence, that the implementation of quality assurance systems in neurologic and trauma rehabilitation is able to improve patients' functional health status is low. Indicators for outcome measurement should be assessed- not only for quality assurance reasons. Neurologic rehabilitation aims at improving the functional health status of patients as well as reintegrating patients in social participation. Therefore a variety of instruments for assessing functional health outcomes, activity and participation of patients have been developed and validated. The comprehensive and multi-dimensional assessment of patients' health status is the basis of allocation decisions, standardised planning in rehabilitation as well as coming up to the patients' and the health care providers' expectations. The selection of appropriate measuring instruments makes significant contributions to improved patient care and its treatment outcomes.

The first part of this project aims at identifying instruments for classifying disease severity in patients with neurologic or trauma rehabilitation and assessing them according to test quality criteria. As an example, chose two specific diagnostic groups (stroke and traumatic brain injury)

We identified 2527 publications by systematic literature search and hand search. 167 full text articles met our inclusion criteria and were included in further analyses.

In stroke patients, specific instruments for classifying the severity of the disease show better test quality than generic instruments. We recommend the National Institut of Health Stroke Scale (NIHSS), the Beck Depression Inventory (BDI), the Frenchay Activities Index (FAI), the Stroke Impact Scale (SIS) and the Stroke Specific Quality of Life Scale (SSQOL) according to test quality criteria. In patients with traumatic brain injury specific instruments do not necessarily perform better than generic instruments. We recommend the Disability Rating Scale (DRS) and the Community Integration Questionnaire (CIQ).

Instrument selection is not only dependant of test quality (and feasibility and acceptability issues), but also of the purpose the measurements should serve in the system. Generic instruments can be used across different diseases, while additionally other criteria have to be assessed disease specific (modular composition of generic and disease specific instruments).

For a choice of measurement instruments in neurological and trauma rehabilitation in Austria, aims of measurements have to be defined. It has to be clarified, whether and in which way measurements for quality assurance and outcome evaluation will be implemented and if it will be connected to reimbursement.

Comprehensive and multi-dimensional assessment of patients' health status are basis of allocation decision and standardised rehabilitation planning

Aim: identifying instruments for classifying disease severity, assessing test quality

167 full text articles included

Stroke: disease specific instruments show better test quality than generic ones

Brain injury: specific instruments do not necessarily perform better

Modular composition of generic and specific instruments

Aims of measurement must be defined

1 Einleitung

Die Notwendigkeit, die Behandlungsergebnisse von Patient/innen nach Schlaganfällen, Unfällen bzw. bei verschiedenen neurologischen Diagnosen zu verfolgen und zu dokumentieren, gewinnt zunehmend an Bedeutung. Zu Qualitätssicherungszwecken gibt es zwar eine Vielzahl von institutionalisierten Akkreditierungs- und Zertifizierungssystemen, die Evidenz aber, dass die Implementierung von solchen struktur- und prozessgesteuerten Systemen eine Verbesserung der funktionalen Gesundheit der Patient/innen mit sich bringt, ist begrenzt.¹ Diese Erkenntnis legt nahe, auch zum Zwecke des Qualitätsmanagements nach Indikatoren zu suchen, die eine ergebnisorientierte Messung der funktionalen Gesundheit ermöglichen. Die funktionale Gesundheit ist ein zentraler Begriff in der Rehabilitation, welcher den Prinzipien des Krankheitsfolgemodells der Weltgesundheitsorganisation (WHO) folgt. Gemäß Internationaler Klassifikation für Funktionsfähigkeit, Behinderung und Gesundheit (International Classification of Functioning, Disability and Health – ICF), sind Gesundheitszustände mehrdimensional. Also sind neben Körperstrukturen und -funktionen auch Aktivität und Partizipation von Bedeutung.

Aufgrund der großen Variabilität der Verletzungen und Krankheitsbilder in der neurologischen Rehabilitation ist die eindimensionale Beurteilung der funktionalen Gesundheit problematisch. Da es aber das erklärte Ziel in der Rehabilitation ist, nicht nur körperliche Funktionen und Organfunktionen wiederherzustellen, sondern den Patienten/innen auch die gesellschaftliche Teilhabe wieder zu ermöglichen, reicht es nicht, die Ergebnisevaluation auf der Ebene der Struktur und Funktion durchzuführen, obwohl diese Art der Ergebnismessung, etwa in der Unfallchirurgie, eine lange Tradition hat.² Neben den einfachen Funktions-Scores wurde eine Vielzahl von Instrumenten zur Beurteilung der Aktivität und der Partizipation entwickelt und validiert.

Eine umfangreiche Beurteilung des Gesundheitszustands setzt eine multidimensionale Evaluation auf mehreren ICF-Ebenen voraus. Erst dieses multidimensionale Assessment bildet eine ausreichende Grundlage, um optimierende Allokationsentscheidungen in der Rehabilitation zu treffen³, eine standardisierte Rehabilitationsplanung zu ermöglichen⁴ und um den Erwartungen der Patient/innen und Kostenträger an das Ergebnis der Rehabilitation gerecht zu werden.⁵ Eine sorgfältige Auswahl von Messinstrumenten und deren Implementierung kann Patientenversorgung und Behandlungsergebnisse verbessern.

**Dokumentation der
Behandlungsergebnisse
zur Qualitätssicherung**

Ergebnisorientierung als

**multidimensionales
Assessment auf Basis
WHO-ICF**

**ermöglicht Aussagen zur
Planung,
Ressourcenallokation,
Ergebnisbeurteilung**

1.1 Ziel und Wissenschaftliche Fragestellungen

Ziel des Projekts ist es

- a. eine systematische Literaturübersicht über international angewandte Methoden und Modelle der Schweregraddifferenzierung und Ergebnisbeurteilung in der Neuro- und Traumarehabilitation zu geben, welche
- b. die identifizierten Messinstrumente nicht nur im Hinblick auf ihre klinische Relevanz (Testgüte, Prognose) bewertet, sondern

**Projekt in drei
Teilbereiche gegliedert**

- c. auch im Hinblick auf ihre Verwertbarkeit zur Performanz-Evaluation der Versorgungsstrukturen betrachtet.
- d. Aus internationalen Erfahrungen mit der Anwendung von Instrumenten zur Schweregraddifferenzierung in der Qualitätssicherung und Ressourcenallokation sollen Umsetzungshinweise in Österreich abgeleitet werden können.

**1. Teil exemplarisch:
indikationsbezogene
Messinstrumente für 2
Diagnoseruppen**

Zur Bearbeitung des Themenfeldes „Neuro- und Traumarehab: Methoden der Schweregraddifferenzierung“ wurde das Projekt in drei Teilbereiche gegliedert:

1. Darstellung von Messinstrumenten zur Schweregrad- und Ergebnisbewertung in der Neuro- und Traumarehabilitation anhand zweier ausgewählter Diagnosegruppen (Schädel-Hirn-Trauma, Schlaganfall) und Bewertung der Testgüte der Instrumente. Es soll die Komplexität der Aufgabenstellung herausgearbeitet und auf die Probleme bei einer indikationsbezogenen Aufbereitung des Themas eingegangen werden.

✿ Welche Instrumente werden zur Schweregraddifferenzierung und Ergebnismessung in der Rehabilitation von Patient/innen mit Schlaganfällen und Schädel-Hirn-Traumata eingesetzt?

✿ Wie sind diese Instrumente im Hinblick auf ihre Testgüte (Sensitivität, Spezifität, prädiktive Werte, Veränderungssensitivität) zu bewerten?

**2. Teil: Internationale
Erfahrungen mit
diagnoseübergreifenden
Messinstrumenten bei
Klassifizierung**

2. Internationale Erfahrungen mit der Anwendung diagnoseübergreifender (sog. generischer) Instrumente zur Schweregraddifferenzierung im Bereich Neuro- und Traumarehabilitation mit Schwerpunkt auf Klassifizierung der PatientInnen und der Verwendung von Klassifizierungen zur Bewertung der Qualität, zur Refundierung, zu Erfahrungen mit Anreizmechanismen.

Welche internationalen Erfahrungen bestehen in der Anwendung der Schweregraddifferenzierung mittels indikationsgruppenübergreifender (generischer) Instrumente

1. zur Qualitätssicherung,
2. als Refundierungsmechanismus
3. zur Performanzevaluierung

**3. Teil: detaillierte
Beschreibung einiger
ausgewählter
Instrumente**

3. Beschreibung ausgewählter indikationsübergreifender Instrumente der Schweregraddifferenzierung in der Neuro- und Traumarehabilitation und Bewertung ihrer Testgüte anhand der in Teil 1 herangezogenen Parameter. Die Systematik der Auswahl ergibt sich aus den vorangegangenen Projektteilen.

**Beurteilung nach Test-
Gütekriterien soll
Auswahl geeigneter
Instrumente erleichtern**

In dem vorliegenden 1. Teil des Reviews werden die international meist verbreiteten Instrumente bei zwei Diagnosegruppen (Schlaganfall und Schädel-Hirn-Trauma) hinsichtlich der Test-Gütekriterien und Einsetzbarkeit in der Rehabilitation verglichen. Es wurden gezielt Publikationen gesammelt, die über Reliabilität und Validität der Instrumente sowie über deren Änderungssensitivität Auskunft geben.

Aus fachlicher Sicht ist es kaum möglich einen starren Satz von Instrumenten zu empfehlen, sondern es braucht einen kombinierten Einsatz von generischen und diagnosespezifischen Instrumenten sowie eine Abstimmung auf die im Gesundheitswesen tätigen Berufsgruppen. Tests mit evaluierten Gütekriterien sollen nicht als abstrakte Größen hingenommen werden, sondern zur bewussten Auswahl von Messinstrumenten unter Berücksichtigung des Patient/innen-Kollektivs, der Diagnose und der Art und Dauer der Rehabilitation genutzt werden.

2 Methodik zur Beurteilung der Messinstrumente

2.1 Messtheorie

Die epidemiologische Forschung unterscheidet eine Vielzahl von Testverfahren, die auch in der neurologischen Rehabilitation angewendet werden. Die Validität von biomedizinischen Testverfahren ist als die Fähigkeit eines Tests, zwischen „Kranken“ bzw. Personen mit einem bestimmten Gesundheitszustand und „Gesunden“ bzw. Personen, bei denen dieser Gesundheitszustand nicht vorliegt, zu unterscheiden, definiert.

Testverfahren müssen zwischen Kranken & Gesunden unterscheiden können

Gemäß der Breite der in der Neurorehabilitation zu messenden Dimensionen des Gesundheitszustands (vgl. WHO/ICF), unterliegt die qualitative Beurteilung der Testverfahren einer Reihe von systematisch anzuwendenden Kriterien, die in diesem Abschnitt erläutert werden.

2.1.1 Validität

Die Validität ist von zwei Komponenten abhängig: der Sensitivität (Wahrscheinlichkeit für ein richtig positives Testresultat) und der Spezifität (Wahrscheinlichkeit für ein richtig negatives Testresultat). Diese Kennwerte können in sogenannte Validationsstudien nur dann bestätigt werden, wenn die tatsächliche Verteilung der abhängigen Variable in der Zielbevölkerung bekannt ist. Bei biophysiological Parametern ist dies weitgehend möglich. In der neurologischen Rehabilitation hingegen ist ein ausschließliches Messen der biophysiological Faktoren nicht hinreichend, weil die Interventionen auf eine Verbesserung des Gesundheitszustands – auch in seinen psychologischen und sozialen Dimensionen – abzielen. Alle Instrumente, die psychologische, soziologische und soziale Faktoren des Genesungsprozesses messen, können nicht anhand der Testgütekriterien für biophysiological Instrumente beurteilt werden, da die tatsächliche Verteilung der Kondition innerhalb der getesteten Population nicht bekannt ist. Die Messinstrumente werden deshalb aus psychometrischer Perspektive konstruiert und beurteilt, was methodisch schwierig ist, da dabei auf eine Reihe von in sich zusammenhängenden Hypothesen der Psychometrie Rücksicht genommen werden muss.

Validität: Befähigung des Tests, Populationen zu unterscheiden

Validitätskriterien biophysiological Tests:

Sensitivität:
Wahrscheinlichkeit für richtig positives Ergebnis

Spezifität:
Wahrscheinlichkeit für richtig negatives Ergebnis

Validitätskriterien psychometrischer Tests

Ein erstes dieser Kriterien ist, ob der Inhalt des zu beurteilenden Messinstruments augenscheinlich dazu geeignet ist, das Richtige zu messen (*face validity*, *content validity*), was i. d. R. von einzelnen oder mehreren Experten/innen im Konsens festgestellt wird. Bei der Augenscheinvalidität handelt es sich um ein streitbares Testgütekriterium, da Expertenmeinungen sich auch erheblich unterscheiden können, wenn es an empirischem Methodenverständnis mangelt.

face /content validity:
wird das Richtige gemessen?

Die Psychometrie stützt ihre Aussagen deshalb auf empirisch gesicherte Informationen.

empirisch abgesichert

2.1.2 Reliabilität

Reliabilität: Reproduzierbarkeit auch unter geänderten Bedingungen	Die Reliabilität ist ein Testgütekriterium, welches das Ausmaß sowohl an systematischen als auch zufälligen Fehlern beschreibt, welches jeder Messung inhärent ist. Während sich das Konzept der Reliabilität einer breiten wissenschaftlichen Zustimmung erfreut, ist es aber in der biomedizinischen Forschung umstritten und unterreflektiert. Dies rührt aus dem Festhalten an der naturwissenschaftlichen Methodologie der Beurteilung der Validität von Tests aufgrund von Sensitivität, Spezifität und prädiktiven Werten. Ebenso wie diese beschreibt die Reliabilität die Fähigkeit eines Tests, eine Testpopulation richtig zu kategorisieren und diese Eigenschaft unter veränderten Bedingungen beizubehalten.
Reliabilitätskriterien: innere Konsistenz	Die innere Konsistenz des Instruments kann anhand von Messungen zu einem bestimmten Zeitpunkt beurteilt werden. Man vergleicht, wie stark die Ergebnisse einzelner Subskalen des Instruments zueinander in Beziehung stehen und berechnet statistische Maßzahlen (Cronbachs alpha, Kuder-Richardson usw.).
Stabilität: Test-Retest-/ Interrater-/Intrarater- Reliabilität	Die Stabilität von Messungen kann zu unterschiedlichen Zeitpunkten (Test-Retest-Reliabilität), von unterschiedlichen Personen (Inter-Observer-Reliabilität/ Interrater Reliabilität) bzw. von denselben Personen zu unterschiedlichen Zeitpunkten (Intra-Observer-Reliabilität) durchgeführt werden und variieren. Die Ergebnisse, die Patient/innen auf den Instrumenten erzielen, variieren aufgrund einer Vielzahl von Faktoren, etwa der Anzahl der zu Messungen eingesetzten Items, der Verteilung sozio-demographischer Charakteristika in der Studienpopulation, systematischer Unterschiede der Patient/innenkollektive und auch, wenn Tests zu unterschiedlichen Zeitpunkten von unterschiedlichen Personen durchgeführt werden.
Vergleich mit „Gold Standard“:	Während die Reliabilität darüber informiert, ob die Testergebnisse reproduzierbar sind, sagt sie nichts darüber aus, was gemessen wurde. In der Regel werden die Testverfahren deshalb mit den bereits etablierten Verfahren verglichen (konkurrente Validität, konvergente Validität). Dabei gelten die bestehenden, etablierten Skalen, mit denen ein neu entwickeltes Instrument verglichen wird als „Gold Standard“.
Vorbedingung für das eigentliche Mess-Ziel der Neurorehabilitation ist das Potential der Tests Änderungen zu messen	Die statistische Beurteilung der Testverfahren ist ein notwendiger, aber nicht ausreichender Schritt, um die Fragestellungen in diesem Review hinreichend zu beantworten. Neurologische Rehabilitationsmaßnahmen werden durchgeführt, um positive Veränderungen des Gesundheitszustandes von Patient/innen herbeizuführen. Daraus folgt, dass das Messen dieser Veränderungen in den Mittelpunkt der Betrachtung rückt.

2.1.3 Änderungssensitivität

Änderungssensitive Messinstrumente vermögen die Ergebnisse von Behandlungen abzubilden	Die Veränderungssensitivität ist die Fähigkeit eines Instruments, klinisch relevante Veränderungen mit einiger Zuverlässigkeit zu messen. Aus dieser Perspektive geraten Reliabilität und Validität nicht in den Hintergrund, sondern sind viel mehr Voraussetzungen für das Messen von Veränderungen des Gesundheitszustands. Aus messtheoretischer Perspektive ist so möglich, die Ergebnisse der Behandlung abzubilden und damit die Effektivität der klinischen/rehabilitativen Interventionen zu untersuchen.
---	--

Die Ziele der Veränderungsmessung selbst sind:⁶

- ✿ die inter-individuellen Unterschiede im Ausmaß der Veränderung zu bestimmen und Subgruppen zu identifizieren, in denen die Therapie/Rehabilitation effektiver ist als in anderen;
- ✿ die Korrelationen der Veränderung zu bestimmen, um jene Faktoren zu identifizieren, die mit den Veränderungen zusammenhängen.

Wenn die Ergebnisse der Rehabilitationsmaßnahmen bei manchen Patient/innen besser sind, als bei anderen, ist es schwieriger, den Effekt der Maßnahme selbst zu bestimmen. In klinischen Studien wird durch Randomisierung eine zufällige Verteilung dieser Einflussfaktoren angenommen, um prognostische Faktoren der Wirksamkeit einer Therapie bestimmen zu können. In der klinischen Tätigkeit, in der die „Wirksamkeit unter realen Bedingungen“ zu eruieren ist, ist es hingegen wünschenswert, zwischen der Wirksamkeit verschiedener Behandlungsstrategien in Subgruppen von Patient/innen unterscheiden zu können, um Aussagen über die Durchführungsqualität (Performanz), der medizinischen und rehabilitativen Leistungen treffen zu können.

Das methodische Vorgehen zur Bestimmung der Änderungssensitivität kann zweierlei sein: Man unterscheidet im Vorhinein zwei Gruppen von Patient/innen anhand des Outcomes, unter Betrachtung der Wahrscheinlichkeit mit der ein festzulegender Gesundheitszustand eintritt, voneinander — und berechnet ausgehend davon die Effektstärke nach Guyatt⁷ oder die „Receiver Operator Characteristic (ROC) Curve nach Deyo and Centor⁸, oder es werden Veränderungen innerhalb desselben Patient/innen-Kollektivs anhand der Effektstärke nach Kazis et al. beobachtet⁹, bzw. wird nach statistisch signifikanten Veränderungen etwa anhand eines t-Tests und anderen Methoden geprüft.

Die Änderungssensitivität eines Instruments oder Mess-Verfahrens ist in der Praxis von vielen Umweltbedingungen abhängig, die zu systematischen Fehlern in der Messung führen können. Würde man beispielsweise versuchen, die Patient/innen direkt nach den Verbesserungen durch die Behandlung zu fragen, wie dies im klinischen Kontext Gang und Gäbe ist („Hat sich Ihr Befinden verbessert, seit Sie das neue Medikament einnehmen? Wie sehr?“), erhielte man ein schwer einschätzbares Ausmaß an verfälschten Antworten. Menschen neigen dazu, sich an den Ausgangs-Gesundheitszustand nur schwer erinnern zu können und machen deshalb Angaben, die stark mit dem aktuellen Gesundheitszustand korrelieren. Eine Reihe von psychometrischen Studien¹⁰ ist der Frage nachgegangen, wie dieser Effekt die Ergebnisse verzerren kann. Etwa:

- ✿ Die Patient/innen unterschätzen den Gesundheits-Ausgangszustand systematisch, was tendenziell zu einer Überschätzung des Therapieeffekts führt.
- ✿ Die fälschliche Einschätzung des Ausgangszustands korreliert mit dem Gesundheitszustand zum Befragungstag (Cluster Bias)

Ziele:

**inter-individuelle
Unterschiede identifizieren**

**Faktoren, die
Veränderungen
beeinflussen identifizieren**

**Aussagen über
Durchführungsqualität**

**verzerrende Faktoren:
systematische Fehler...**

**...wirken sich auf die
Ergebnisse aus**

**Überschätzung des
Therapieeffekts bei
Unterschätzung des
Ausgangszustands**

weilers:	Darüber hinaus sind unter den „realen“ Bedingungen der rehabilitativen Behandlungen auch die spezifische Zusammenstellung des Patient/innen-Kollektivs, insbesondere die Verteilung von Alter und Schweregrad (sog. „Case mix“) als klassische Verzerrungsfaktoren zu berücksichtigen. Patient/innen könnten zudem aus weiteren Gründen (Befragungssituation, soziale Erwünschtheit) dazu neigen, einen insgesamt besseren oder auch schlechteren Gesundheitszustand als den tatsächlichen anzugeben.
Case-Mix	
Einflüsse durch Befragungssituation; Antizipation von sozialer Erwünschtheit	
ÄrztInnen tendieren zur Überschätzung des Gesundheitszustandes im Vergleich zur Selbstbewertung oder durch Angehörige	Die Aufgabe für die Testkonstruktion und -administration besteht darin, diese und weitere systematische Verzerrungen weitestgehend zu reduzieren. Wenn der Gesundheitszustand z.B. vom Gesundheitspersonal und nicht von Patient/innen selbst und/oder von Angehörigen beurteilt wird, kommt es in der Regel auch zu unterschiedlichen Verzerrungen. In der Regel überschätzen Ärzte/innen den Gesundheitszustand zu Beginn einer Behandlung und überschätzen ihn bei Entlassung, wenn man ihre Bewertungen mit denen Anderer in der neurologischen Rehabilitation tätigen Gesundheitsberufen, den Selbstbewertungen der Patient/innen oder Bewertungen durch Angehörige vergleicht.
Ceiling- & Floor-Effekte machen die Messung von Veränderungen schwerer	Ein weiteres Problem der Änderungssensitivität, das in der Literatur spezifisch reflektiert wird, sind sog. Floor- und Ceiling-Effekte, die dann entstehen, wenn ein Teil der evaluierten Einzelpersonen ihren Gesundheitszustand auf der Skala sehr schlecht oder sehr gut bewerten. Für die Messung von Veränderungen sind diese effekte relevant, wenn der Ausgangszustand von weniger schwer erkrankten (oder verunfallten) Patient/innen sehr gut ist, da positive Veränderungen nicht gemessen werden können, oder umgekehrt Verschlechterungen bei niedrigem Ausgangsniveau nicht messbar sind.

2.2 Operationalisierung der Evaluationskriterien

Kontextfaktoren erschweren die Operationalisierung:	Salter et al. ¹¹ bereiteten das bis dahin vorhandene Wissen zur Frage, wie Messinstrumente auf Basis ihrer Testgütekriterien klassifiziert und in der Rehabilitation eingesetzt werden können, auf. Die Glaubwürdigkeit des Evaluationsprozesses hängt nicht nur mit Reliabilität, Validität und Änderungssensitivität der Instrumente zusammen, sondern auch mit Präzision, Interpretierbarkeit, Akzeptierbarkeit und Machbarkeit. So sind z.B. das Vertrauen in Evaluationsergebnisse mit dem Verwaltungsaufwand für die Evaluation assoziiert. ¹² Dijkers et al. ¹³ stellen fest, dass diese kontextabhängigen Evaluationskriterien in bemerkenswert wenigen Evaluationsstudien zur Rehabilitation von den Autor/innen, zur Verteidigung der gewählten Methodik angeführt oder zur Diskussion der Ergebnisse, verwendet werden.
Interpretierbarkeit, Akzeptierbarkeit, Machbarkeit	
Bedürfnisse der AnwenderInnen beachten	Während viele Studien auf die Bewertung der klinischen Wirksamkeit rehabilitativer Interventionen abzielen, fehlt meistens Konsens darüber, welche Test- und Messinstrumenten zur Evaluation angewendet werden sollen. In der Praxis der Rehabilitation kommt hinzu, dass diese Instrumente auf die Bedürfnisse einer Vielzahl von Betroffenen (Patient/innen, Leistungserbringer/innen, Entscheidungsträger) abgestimmt sein müssen. Für eine aussagekräftige Evaluierung der Rehabilitation ist es notwendig, ein verbindliches System der Ergebnismessung einzuführen. ¹⁴
Rahmenkonzept ICF (WHO)	

Dafür bietet das ursprünglich 1980 von der WHO veröffentlichte Konzept der International Classification of Functioning, Disability and Health (ICF) einen multidimensionalen Rahmen.^{15 16} In seiner aktuellen Version identifiziert das ICF Rahmenkonzept drei Ebenen menschlicher Funktionalität: eine bio-physiologische Ebene (früher: „impairment“; heute: „body function/structure“), eine an den Aktivitäten orientierte (früher: „disability“; heute: „activity“) sowie eine zur Interaktion und Teilhabe Einzelner mit/an ihrer Umwelt (früher: „handicap“; heute: „participation“). Aus der Sicht der Evaluationsforschung müssen diese Ebenen als abfallend betrachtet werden, weil es bei „Aktivitäten“ und „Teilhabe“ zunehmend schwieriger wird, gemessene Veränderungen auf rehabilitative Interventionen zurückzuführen, da sie alternativ als Folge einer Reihe von nicht erfassten/erfassbaren Faktoren entstanden sein könnten.¹⁷ Steiner & Norman zufolge bleibt unklar, ob die drei Ebenen der ICF eindeutig und ausschließlich genug sind, um in der systematischen Evaluationsforschung bestehen zu können.¹⁸ Das Rahmenkonzept der ICF wurde von ihnen vielmehr als Beitrag für eine Klassifizierung der Messinstrumente interpretiert, um den internationalen fachlichen Austausch zum Thema zu vereinheitlichen.

**body function /
structure
activity
participation**

Während Fitzpatrick et al.¹⁹ Guidelines für die Anwendung von Messinstrumenten in klinischen Studien entwarfen, fehlen diese für den Einsatz in der rehabilitativen Versorgung. Nichtsdestotrotz verwenden Salter et al.²⁰ die ICF und operationalisieren acht Evaluationskriterien der Testgüte samt quantifizierbaren Standards für Reliabilität, Validität und Änderungssensitivität.

**Guidelines oder
Standards zur
Anwendung von
Messinstrumenten in
der Rehabilitation
fehlen**

Tab. 2.2-1: Übersicht zur Operationalisierung der Evaluationskriterien

KRITERIUM	DEFINITION	STANDARD
Zweckdienlichkeit/ Sinnhaftigkeit	Übereinstimmung mit Ziel und Fragestellung	kontextabhängig
Reliabilität	Wiederholbarkeit und interne Konsistenz anhand von -Test-Retest Reliabilität -Reliabilität von einzelnen Items -Items zu Skalen - einzelner Beobachter (Intrarater) - zwischen Beobachtern (Interrater)	Test Retest bzw. Interrater Reliabilität (ICC; kappa statistics): Ausgezeichnet $\geq 0,75$ Adäquat $0,4 - 0,74$ Mangelhaft $\leq 0,4$ Internal consistency: Ausgezeichnet: $\geq 0,8$ Adäquat: $0,7 - 0,79$ Mangelhaft: $\leq 0,7$ Inter-item und Item-to-scale Korrelationskoeffizienten: Inter-item: Adäquat: $0,3 - 0,9$ Item-to-scale: Adäquat: $0,2 - 0,9$
Validität	Augenscheinvalidität, Inhaltsvalidität, Konstrukt- und Kriterienvvalidität (concurrent, convergent, und predictive)	konvergente und konkurrente Konstruktvalidität: Ausgezeichnet: $\geq 0,6$ Adäquat: $0,31 - 0,59$ Mangelhaft: $\leq 0,3$ ROC-Analyse – AUC: Ausgezeichnet: $\geq 0,9$ Adäquat: $0,7 - 0,89$ Mangelhaft: $\leq 0,7$
Änderungssensitivität	Veränderungen bei Patient/innen zwischen zwei Messzeitpunkten (therapeutische Effekte); Bias durch „floor“ oder „ceiling“ Effekte	Standardized effect size: groß: $\geq 0,8$ moderat: $0,5 - 0,8$ klein: $\leq 0,5$ Standardized response mean: Mangelhaft: schwache Evidenz nur auf Basis von p-Werten (stat. Sign.) Floor/Ceiling Effekte: Ausgezeichnet: keine Adäquat: $\leq 20\%$ Mangelhaft: $> 20\%$
Präzision	Anzahl der Merkmalsausprägungen bei Ordinaldaten	abhängig von Zielen und Fragestellung
Interpretierbarkeit	Aussagekraft der Punktwerte/Indizes; terminologische Klarheit; Vergleichbarkeit der Ergebnisse;	kontextabhängig; z.B. problematisches Handling; Bias und Confounding
Akzeptierbarkeit	Repräsentation einer Krankheitslast; Kann der Einfluss (von Dritten) geschätzt werden?	kontextabhängig; z.B. fehlende Werte; Bias
Machbarkeit	Ausmaß des organisatorischen Aufwands, Ausgaben und Störung des therapeutischen Ablaufs durch Administration des Messinstruments	kontextabhängig

3 Methodik: Systematische Literaturrecherche

Um Publikationen zur Testgüte der einzelnen in der neurologischen Rehabilitation eingesetzten Messinstrumenten systematisch zu suchen, wurden die seit 1995 veröffentlichten wissenschaftlichen Arbeiten in Peer Reviewed Journals einer systematischen Bewertung mit den im Folgenden ausführlich dargestellten Ergebnissen unterzogen. Die Literatursuche wurde in den Datenbanken Medline, Embase, Cochrane und in einschlägigen HTA-Datenbanken durchgeführt.

**Medline, Embase,
Cochrane und
einschlägige HTA-
Datenbanken 1995-2008**

3.1 PICO-Frage und Suchstrategie

Tab. 3.1-1: PICO-Suchstrategie

Population	Patient/innen mit den Diagnosen Schlaganfall, Schädel-Hirn-Trauma, Morbus Parkinson, Apallisches Syndrom, Multiple Sklerose, Demenz, Dystonie sowie neuromuskuläre Erkrankungen
Intervention	Assessment der Ergebnisqualität Performanz-Evaluation
Kontroll- intervention	keine
Outcomes	Funktionalität, Aktivitäten des täglichen Lebens, gesundheitsbezogene Lebensqualität, Patient/innen-Zufriedenheit Effektivität, Effizienz
Setting	USA, CAN, AUS, NZL, Westeuropa
Studiendesigns	keine Einschränkung

3.2 Literatursuche

Literatursuche: Insgesamt wurden 2509 (Medline 1389; Embase 1009; Cochrane 141) Zitate in der systematischen Literatursuche sowie 18 weitere durch manuelle Suche gefunden, was auf 2527 summierte.

2527 Zitate

Der Auswahlprozess ist in Abb. 3.2-1 dargestellt:

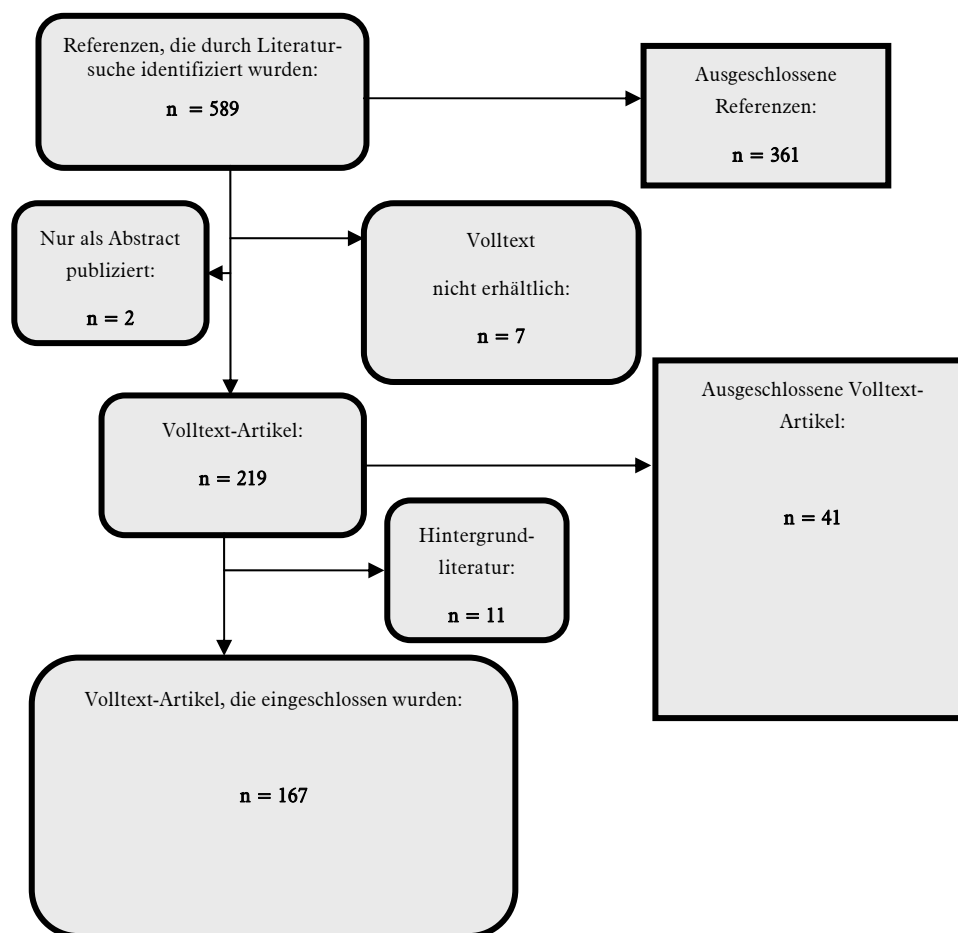


Abb. 3.2-1: Darstellung des Auswahlprozesses (QUORUM tree)

Ausschlusskriterien: Ausgeschlossen wurden all jene klinischen Studien, in denen zwar entsprechende Evaluationsinstrumente eingesetzt wurden, die aber nicht auf deren Evaluation zur Bestimmung von Testgütekriterien fokussiert waren (Wirksamkeitsforschung). Weiters wurden Studien als zu spezifisch betrachtet und ausgeschlossen, wenn sie sich ausschließlich auf Messungen biophysiologicaler Parameter, wie etwa in der medizinisch-technischen Diagnostik, fokussierten. Auf diese Art und Weise wurden etwa 68 % der Treffer ausgeschlossen.

4 Schlaganfall

4.1 Schweregraddifferenzierung

Im Folgenden werden die Ergebnisse der Literatursauswertung systematisch dargestellt. Die Auflistung beginnt bei Instrumenten zur Schweregraddifferenzierung und arbeitet sich in der ICF-Klassifikation Dimension für Dimension nach außen (Aktivität, Partizipation) vor. Dadurch verändert sich der Fokus von exklusiven Schweregraddifferenzierungsskalen und bezieht immer mehr Ergebniskategorien der Neurorehabilitation ein, bis hin zu diagnosespezifischen Skalen zur Erhebung der gesundheitsbezogenen Lebensqualität (Abb. 4.1-1).

Abfolge des Berichts:
von Messinstrumenten
zur physiologischen
Schweregraddifferenzierung
zur
gesundheitsbezogenen
Tests

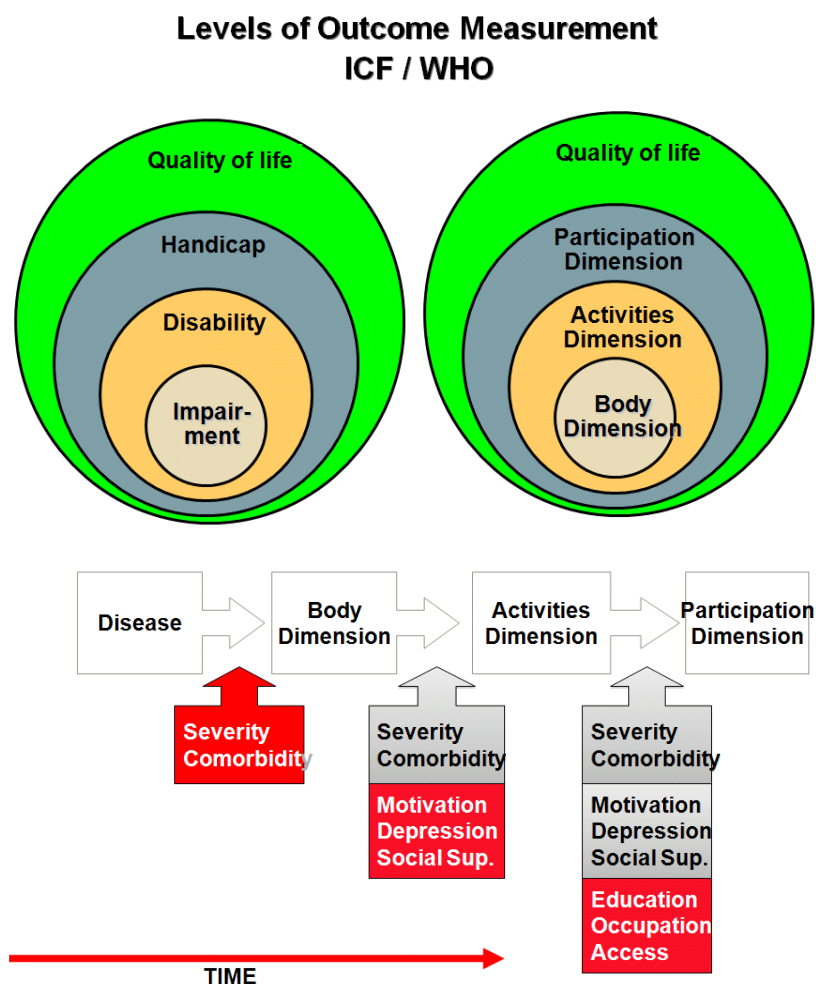


Abb. 4.1-1: Darstellung der Ergebnisse von der Schweregraddifferenzierung zur Ergebnismessung: entlang der ICF-Dimensionen

Aus dem ICF-Modell wird ersichtlich, dass der prognostischen Vorhersagekraft jener Skalen, die nur in der Körper/Struktur-Dimension messen, auf alle zeitlich nachgelagerten weiteren Dimensionen des Modells in der Literatur besondere Aufmerksamkeit gewidmet wird. Sollte der Fortschritt von Genesungsprozessen, gemessen anhand der Dimensionen Aktivität und Partizipation mit einiger Exaktheit vorhersagbar sein, ließe sich ein Messsystem konstruieren, welches ausschließlich auf die Körper/Struktur-Dimension beschränkt bleiben kann und dennoch valide Messungen erzielt.

4.1.1 National Institute of Health Stroke Scale (NIHSS)

**Schweregrad:
Beurteilung des
neurologischen Status,**

15 Items

**Ergebnis: 0-42 Punkte
zu Kategorien
zusammengefasst**

10-15 min

**umstrittene Interrater
Reliabilität,**

**aber im akzeptablen
Bereich**

**auch retrospektive
NIHSS-Bewertung aus
Patientenakte möglich**

**Prognose negativer
Outcomes mit NIHSS ist
möglich, nicht aber
positiver Outcomes**

Die NIHSS ist eine Schlaganfall-spezifische Skala zur Bestimmung des Schweregrades. Sie besteht aus 15 Items, die sehr eng an den neurologischen Status angelehnt sind und in einem Ergebnisbereich von 0-42 Punkten resultieren. Je höher die Punktzahl, desto schwerer die Folgen des Schlaganfalls. Die Beurteilung lässt eine Zuordnung in sechs unterschiedlichen Kategorien zu (Bewusstsein/Aufmerksamkeit, Visus, Motorik, Koordination, Sensibilität, Sprache). Die Anwendung der Skala erfordert im Allgemeinen ein spezifisches, Video-unterstütztes Training und bei der Anwendung zu Studienzwecken zusätzlich eine Zertifizierung. Eine Bewertung auf der Skala dauert 10-15 Minuten.

Josephson et al.²¹ testeten die NIHSS mit einer großen Anzahl von Kliniker/innen. 7405 Ärzt/innen und Pflegende beurteilten die elf, zwischen 1998 und 2004 für das Training empfohlenen, Videovignetten von Patient/innen. Bei 7 dieser 11 Patient/innen-Videos traten signifikante Unterschiede in der Bewertung (mehr als 4 Punkte im Summenwert) auf. Die Interrater Reliabilität, die bis dahin schon als Schwäche der Skala gegolten hatte²², bestätigte sich auch in der aktuellsten Studie zum Thema. Meyer et al.²³ evaluierten eine modifizierte Form der NIHSS, die aus weniger Items bestand, und fanden sehr ähnliche Ergebnisse.

In einer portugiesischen Sprachvalidationsstudie mit 51 Schlaganfall-Patient/innen erzielte die NIHSS hingegen bessere Interrater Reliabilitätswerte als die modified Rankin Scale/mRS und der Barthel Index/BI.²⁴ Berger et al.²⁵ übersetzten die NIHSS, die European Stroke Scale/ESS und die modified Rankin Scale ins Deutsche und testeten ihre Interrater Reliabilität in einer Studie mit 43 Patient/innen. Die NIHSS erreichte eine Reliabilität von $r=0,8$, die European Stroke Scale $r=0,79$ und die Rankin Scale $r=0,76$. Alle diese Werte liegen in einem akzeptablen statistischen Testgütebereich. Die Evidenz zur Interrater Reliabilität ist also umstritten.

Mehrere Studien gingen der Frage nach, ob die NIHSS-Werte mit Patient/innen-Daten aus Krankenakten rekonstruiert werden können. Diese Frage ist etwa für Studien relevant, in denen der Schweregrad retrospektiv eingeschätzt werden soll. Kasner et al. fanden heraus, dass die direkten Bewertungen auf der Skala mit den aus Patient/innendaten zufriedenstellend korrelierten ($r=0,79-0,85$).²⁶ In einer anderen Studie konnten aus Krankenakten Patient/innen mit gutem Gesundheitszustand mit einer Sensitivität von $r=0,72-0,89$ identifiziert werden.²⁷ Williams et al.²⁸ entwickelten in dieser Frage sogar einen Algorithmus für fehlende Daten, da in ihrer Studie nur die Daten von ein/er/m von 32 Patient/innen die Informationsdichte lieferten, die notwendig gewesen wäre, um die NIHSS vollständig ausgefüllt

zur Verfügung zu haben. Auch sie kamen zum Schluss, dass eine retrospektive Bestimmung der NIHSS-Werte über Patient/innendaten möglich ist.

Andere Studien haben die Fragestellung bearbeitet, mit welcher Genauigkeit die Schweregraddifferenzierung auf der NIHSS in der Lage ist, Mortalität und Outcome vorherzusagen. Muir et al. hatten die prognostischen Vorzüge der NIHSS gegenüber anderen bis dahin gebräuchlichen Scores bereits 1996 erwiesen.²⁹ In einer Studie mit 359 Patient/innen stellten Cheung et al.³⁰ (2008) fest, dass die NIHSS mit einer Sensitivität von 0,8 und Spezifität von 0,81 in der Lage ist, Mortalität innerhalb von 30 Tagen vorherzusagen, wenn ein Cut-off von 20 überschritten wird. Die 5-Jahres-Mortalität kann mit einer Sensitivität von 0,88 und Spezifität von 0,57 gemessen werden. Diese relativ akzeptablen Werte verschlechtern sich bei der Prognose der Funktionalität („Aktivität“). Die Sensitivität liegt hier bei inakzeptablen 0,41.

4.1.2 Die Orpington Prognostic Scale (OPS)

Bei der Orpington Prognostic Scale handelt es sich um die Weiterentwicklung der Edinburgh Scale und um Items im Bereich der Kognition (Hodkinson's mental test score). Die Skala besteht aus 15 Items in 4 Kategorien (Motorik, Koordination, Balance, Kognition) und kann in etwa 10-15 min komplettiert werden.

15 Items

10-15 min.

Rieck & Moreland evaluierten die OPS in einer Studie mit 94 Patient/innen mit dem Ergebnis ausgezeichneter Reliabilitätswerte (Interrater Reliabilität und Test-Retest-Reliabilität $r > 0,9$).³¹

Studenski et al. berichten in einer methodisch hochwertigen Studie, dass die Baseline-Werte der OPS die Genesungsraten nach sechs Monaten für fünf Kriterien der Funktionalität gut bis sehr gut vorhersagen, darunter Pflege(un)abhängigkeit, selbständiges Zubereiten von Mahlzeiten, Mobilität, Selbstversorgung mit Medikamenten.³² Shoemaker et al. zeigten, dass die Werte auf der OPS auch mit dem FIM korrelieren, allerdings anhand einer kleinen Patient/innengruppe von $n=22$.³³

Celik et al.³⁴ verglichen die Prognosekraft der NIHSS mit jener der Orpington Prognostic Scale und zeigten, dass beide Skalen die Funktionalität (gemessen anhand des Barthel Index) bis zu einem Monat nach der Diagnose zufriedenstellend, nach 3 und 6 Monaten nur mehr unzufrieden stellend vorherzusagen in der Lage waren, allerdings mit leichten Vorteilen der Orpington Prognostic Scale gegenüber der NIHSS. Wright et al.³⁵ kommen zu sehr ähnlichen Ergebnissen, indem Sie die NIHSS innerhalb der ersten 24 Stunden und die Orpington Prognostic Scale innerhalb einer Woche nach Aufnahme in ein Rehabilitationszentrum verglichen. Bzgl. der Vorhersagekraft beider Skalen bestanden dabei nur geringfügige Unterschiede. In einer aufwändigeren statistischen Analyse nach dem Rasch-Modell erwies sich die NIHSS auch in der Messung geringfügiger Verbesserungen annehmbar sensitiv (0,71).³⁶ In einer Studie, die die Prognosekraft der Edinburgh Scale, Orpington Prognostic Scale und des Barthel Index (IDL) bei über 75jährigen Patient/innen mit Schlaganfall miteinander verglich, schnitt die Orpington Prognostic Scale insbesondere deshalb besser als die anderen ab, weil sie auch bei Demenzerkrankten sensitiv misst.³⁷

**kurzfristige (30 Tage)
Prognose
zufriedenstellend,
langfristig aber nicht**

4.1.3 Die Glasgow Coma Scale (GCS)

mangelhafte Interrater Reliabilität

Die Glasgow Coma Scale wurde ursprünglich entwickelt, um Bewusstsein bei Schädel-Hirn-traumatisierten Patient/innen zu erfassen. Die Skala wird jedoch auch zur Schweregrad differenzierung bei Schlaganfall verwendet. Die Messung basiert auf motorischen und verbalen Fähigkeiten sowie Augenaufschlag und misst Tiefe und Dauer komatöser Gesundheitszustände.³⁸

Wie im Kapitel zu Schädel-Hirn-Traumata im Detail beschrieben, gehört eine mangelhafte Interrater Reliabilität zu den größten Schwächen der Skala³⁹, die jedoch mittels Training des Personals einigermaßen kompensiert werden kann.⁴⁰

Prognose negativer Outcomes (Mortalität) mit GCS ist möglich, nicht aber anderer Outcomes

Weir et al.⁴¹ testeten, inwiefern die auf der Glasgow Coma Scale erzielten Werte mit den körper/strukturbezogenen Outcomes zusammenhängen. Während die Mortalität nach 2 Wochen und der Gesundheitszustand nach 2 Monaten einigermaßen genau vorhergesagt werden konnten (AUC 0,76-0,78), zeigte die Skala Schwächen bei der Vorhersage der Unterbringungsform nach 2 Monaten. Ähnliche Ergebnisse zur Vorhersage der Mortalität hatten bereits auch andere erbracht.⁴² Die Vorhersagbarkeit von Outcomes in den Dimensionen Aktivität und Partizipation wurde bislang nicht geprüft.

Vergleich der Prognosekraft mehrerer Instrumente: alle nicht zufriedenstellend

Counsell et al.⁴³ rechneten 2004 ein statistisches Modell, in dem sie Patient/innen mit einem modified Rankin Score/mRS ≤ 2 drei Monate, sechs Monate und ein Jahr nach dem Schlaganfall mit Vorhersagen verglichen, die auf Alter, einer spezifischen Schweregradklassifikation (Bewusstsein, Inkontinenz) und den Vorhersagen von Arzt/innen beruhten. Die Vorhersagekraft war insgesamt relativ unzufriedenstellend. Anhand einer ROC-Analyse bestätigte sich, dass das Modell auf der Basis der modified Rankin Scale das Outcome immerhin gleich gut und besser vorherzusagen in der Lage ist, als Skalen, die dafür die Krankenakten heranziehen.

4.1.4 Modified Ranking Scale (mRS)

Die Ranking Scale besteht im Original seit 1957 und gilt als weit verbreitete Skala zur Einschätzung des Schweregrades von Schlaganfall-Patient/innen. Sie unterscheidet den Schweregrad in 7 Stufen (0-6) zwischen „symptomfrei“ und „Tod“ und fokussiert auf instrumentelle Aktivitäten des täglichen Lebens, insbesondere motorische Funktionalität. Der Unterschied zwischen der ursprünglichen Version (1957) und der modifizierten Version besteht in einer feineren Differenzierung (eine Stufe mehr) leichter und mittlerer Schweregrade.

Trotz dieser Fokussierung schließt die mRS auch Items ein, die Partizipation als Ergebnisdimension einschließen, wodurch sie sich vom Barthel Index unterscheidet.⁴⁴

Banks et al.⁴⁵ analysierten 50 Studien, die über die Testgütekriterien der mRS zwischen 1957 und 2006 in Peer reviewed Journals veröffentlicht wurden.

Die Test-Retest-Reliabilität wird in 5 Studien adressiert. Sie ist generell hoch und kann durch strukturierte Interviews erhöht werden. Die Interrater Reliabilität ist mitunter nur durchschnittlich. Bei Van Swieten et al.⁴⁶ beträgt sie $r=0,56$ (overall), bei Wolfe et al.⁴⁷ $r=0,75-0,96$, mit signifikanten Unterschieden zwischen beurteilenden Personen. Die Ergebnisse von Wilson et al.⁴⁸ sind sehr ähnlich. Um die Interrater Reliabilität zu erhöhen, gilt es heute als Standard, die Qualität der Datengenerierung durch Administration in strukturierten Interviews zu erhöhen.^{49 50}

Die Kategorien der mRS sind diskriminativ. Den Vergleich mit anderen Instrumenten besteht die mRS konsistent. Die Validität der Skala ist insgesamt akzeptabel belegt, wenngleich auch die Korrelation zu sozialer Partizipation fehlt.⁵¹

Die Vorhersagekraft der mRS ist umstritten. In einer Studie mit 95 Patient/innen mittleren Schweregrades - also jenes Schweregrads, der im Vergleich der Instrumente am schwierigsten sensitiv zu erfassen ist -, schnitt die mRS im Vergleich mit dem Barthel Index (BI) und dem Functional Independence Measure (FIM) schlecht ab. Ebenso ist mit Floor-Effekten von 18 % zu rechnen.⁵²

Nichtsdestotrotz identifizieren Banks et al.⁵³ 25 Studien in verschiedenen Settings, die zeigen, dass die Skalenwerte der mRS mit körperbezogenem intermediärem (schwere Behinderung, Tod) und längerfristigem Outcome korrelieren und diese Eigenschaft unter veränderten Bedingungen (Organisation, Subdiagnosetypen und Schweregrade von Schlaganfall) beibehalten werden kann. Allerdings handelte es sich um in der ambulanten Rehabilitation betreute Patient/innen. Im Akutsetting ist die klinische Sensitivität der mRS nicht so gut belegt.

Andere Schwächen der mRS hängen zum ersten damit zusammen, dass es sich bei der mRS um ein diagnosespezifisches Instrument handelt, welches die oft komorbiden Patient/innen (kardiovaskuläre Erkrankungen, Diabetes, Arthritis) nur hinsichtlich ihrer von Schlaganfall beeinträchtigten Funktionalität erfasst. Außerdem beweisen andere Studien, dass der Gesundheitszustand vor dem Schlaganfall die Werte, die Patient/innen auf der mRS erzielen, beeinflusst.⁵⁴ Weitere Verzerrungsvariablen sind sozioökonomische Faktoren.⁵⁵ Es ist daher äußerst relevant, diese Faktoren in ein zu entwickelndes Messsystem mit aufzunehmen.

4.1.5 Kombinierte Testbatterien

Bei Masiero et al.⁵⁶ kam eine auf mehrere ICF-Dimensionen ausgedehnte Testbatterie zum Einsatz. Untersucht wurden Vorhersagefaktoren für Schlaganfall-Patient/innen, die stationäre Rehabilitation in Anspruch nahmen. Dafür setzten Sie das Functional Independence Measure/FIM, den upper and lower Motricity Index und den Trunk Control Test ein und konstruierten aus diesen Instrumenten den Functional Ambulation Classification Score, den sie bei der Entlassung der Patient/innen erhoben. Wenn sie das Alter der Patient/innen berücksichtigten entstand so ein Instrument, das aufgrund des Alters in Kombination mit motorischer Funktionalität mit einer viel höheren Testgüte als bisher (Sensitivität=86,5 %; Spezifität=85,4 %) in der Lage war, vom Gesundheitsstatus bei der Aufnahme auf den Gesundheitsstatus bei der Entlassung zu schließen. Interessanterweise steigt die Chance, den Schweregrad eines Schlaganfalles exakt zu messen und den

**Interrater Reliabilität
nur durchschnittlich**

**zur Erhöhung der
Datenqualität:
strukturierte Interviews**

Validität akzeptabel

**wenig sensitiv bei
mittleren
Schweregraden: Floor
Effekte**

**Anwendung in
ambulanter Rehab: gute
Vorhersagekraft**

**.... im Akutsetting
umstritten**

**mRS als
diagnosespezifisches
Instrument erfasst nur
Schlaganfall-bedingte
Funktionalität**

**Kombination aus
Instrumenten verbessert
Schweregrad-
einschätzung &
Vorhersagekraft**

<p>Versuch der Kombination von Faktoren zu Ressourcenverbrauch & Schweregrad-Einschätzung zu kombinieren</p>	<p>Gesundheitszustand vorherzusagen, wenn mehrere Dimensionen der ICF bei der Messung berücksichtigt werden.</p> <p>Zieht man wie Huybrechts & Caro⁵⁷ in einem systematischen Review organisations- und steuerungsrelevante Faktoren wie die benötigten Behandlungsressourcen in Abhängigkeit des Schweregrades in Betracht (Barthel Index und modified Rankin Scale kamen zum Einsatz), findet man insgesamt wenig überzeugende Evidenz für beide Skalen, der Barthel Index schneidet noch etwas besser ab als die Rankin Scale. Diesbezüglich werden auch Instrumente, die die ICF-Skala umfassender abbilden, zu bewerten sein.</p>
---	---

4.2 Outcome measures/Ergebnismessung

<p>Cochrane Review : Instrumente zur Ergebnismessung</p>	<p>Roberts et al. untersuchten, welche Outcomes anhand welcher Instrumente in klinischen Studien der Cochrane Stroke Group`s database zwischen 1955 und 1995 verwendet wurden und identifizierten 174 Studien. Eine Kategorisierung entlang der WHO-ICF (alt) ergab, dass 76 % „impairment“, 42 % „disability“, aber nur 2 % „handicap“ oder „quality of life“ als Ergebniskriterien untersucht wurden. Die meisten Studien verwendeten in Bezug auf Inhalt, Reliabilität und Validität inadäquate Instrumente, so die Schlussfolgerung der Autoren.⁵⁸</p>
---	---

4.2.1 Beck Depression Inventory (BDI)

<p>Symptommessung anhand 21 Items</p> <p>hohe Prävalenz bei Schlaganfall-Patient/innen</p>	<p>Die Beck Depression Inventory misst die Intensität depressiver Symptome.⁵⁹ Die Items wurden aus klinischer Erfahrung generiert und bilden keinen spezifischen theoretischen Zugang zu Depression oder ihrer Diagnostik ab. Die Skala besteht aus 21 Items, die Symptome oder Verhalten erfragen, die mit Depression zusammenhängen. Jedes Item wird durch die Zustimmung zu einer von vier Aussagen erfasst, die sich in ihrem Schweregrad unterscheiden. So entsteht ein Gesamtscore von 0-63. Ab einem Summenwert von 10 spricht man in folgenden Abstufungen von einer Depression: 10-18 („mild“), 19-29 („moderate“), 30-63 („severe“).⁶⁰ Die Skala kann als Interview oder als Selbstausfüller-Fragebogen administriert werden. Das Ausfüllen dauert etwa 5-10 min.</p> <p>Aben et al. untersuchten 202 Schlaganfall-Patient/innen ohne Komorbidität innerhalb von 2 Jahren auf die Prävalenz und Messbarkeit von Depression mit verschiedenen Instrumenten.⁶¹ 24 % dieser Patient/innen verweigerten die Teilnahme an der Studie, zudem traten Probleme mit fehlenden Werten auf, was die Prävalenz in Summe wahrscheinlich unterschätzt. Die Prävalenz betrug 25 %, ca. 15 % „major depressive disorder“ und ca. 10 % „minor depressive disorder“ (DSM-IV). Abb. 3 zeigt einen Vergleich der eingesetzten Messinstrumente:</p>
--	---

Measure	Cutoff ^a	Sensitivity		Specificity		Positive Predictive Value		Negative Predictive Value		Area Under the Curve	
		Major Depression Only	Major and Minor Depression	Major Depression Only	Major and Minor Depression	Major Depression Only	Major and Minor Depression	Major Depression Only	Major and Minor Depression	Major Depression Only	Major and Minor Depression
SCL-90 depression subscale	25	88.5	87.8	60.7	65.9	28.0	43.9	96.8	94.7	0.81	0.85
Beck Depression Inventory	10	80.0	77.1	61.4	65.4	22.2	37.5	95.7	91.4	0.78	0.79
Hospital Anxiety and Depression Scale											
Depression subscale	8 ^b	73.1	72.5	81.6	78.9	41.3	50.9	94.5	90.5	0.82	0.83
Anxiety subscale	5	91.7	88.5	56.1	71.8	25.9	63.5	97.7	91.9	0.78	0.77
Total	11	91.7	86.8	65.3	69.9	30.1	45.2	98.0	94.9	0.83	0.84
Hamilton Depression Rating Scale	12	78.1	78.4	74.6	81.3	36.8	58.8	94.7	91.7	0.83	0.86

^aMinimum score at which a possible case of depression is indicated.
^bCutoff = 7 if screening for both major and minor depression.

Abb. 4.2-1: Vergleich der Testgüte verschiedener Depressionsskalen

Die Autoren betonen, dass bis dahin keine Studien, die die Testgüte einzelner Instrumente für Depression bei Schlaganfall-Patient/innen evaluierten, durchgeführt wurden und attestieren allen Instrumenten Ergebnisse, die sie für einen Einsatz eignen. Der Vorteil des BDI liege einer britischen Studie zufolge in seiner guten Akzeptanz, weiten Verbreitung und einfachen Administrierbarkeit.⁶²

Depression als Folge eines Schlaganfalles ist hochprävalent und beeinflusst den Heilungsverlauf negativ. Welches Instrument für die Assessments eingesetzt werden soll, ist von der Akzeptanz in der jeweiligen Situation der Leistungserbringung abhängig.

Validität aller untersuchten Instrumente, BDI hat hohe Akzeptanz

64 Items in 8 Kategorien zu Funktionalität, Aktivität, Partizipation

4.2.2 Stroke Impact Scale Version 3.0 (SIS 3.0)

Die Stroke Impact Scale/SIS ist ein Schlaganfall-spezifisches Messinstrument des Gesundheitsstatus, welches den Anspruch erhebt, Behandlungsergebnisse umfassend im Sinne der ICF-Klassifikation 1-3 abzubilden.⁶³ Die Version 2.0 der Skala besteht aus 64 Items in 8 Kategorien (Kraft, Funktionalität der Hände, Aktivitäten des täglichen Lebens (ADL/IADL), Mobilität, Kommunikation, Emotion, Erinnerung und Denken, Partizipation).⁶⁴ Für die Version 3.0 wurden 5 Items entfernt.⁶⁵ Die SIS beruht auf einer Entwicklung, an der Patient/innen und Behandlungspersonal gleichermaßen teilnahmen. Die Antworten auf die 5er-Likert-Skala beziehen sich auf Schwierigkeiten in den verschiedenen Kategorien in der letzten Woche. Mit einem dem SF-36 vergleichbaren Algorithmus rangieren die Ergebnisse in jeder Kategorie zwischen 0 und 100. Die physiologischen Kategorien können zu einem Summenscore kombiniert werden, alle anderen Kategorien sollen einzeln ausgewertet werden.⁶⁶ Ein extra Item, skaliert in einer visuellen Analogskala fragt nach dem subjektiven Empfinden von Veränderungen. Ursprünglich wurde die Skala als persönliches Interview mit einer Dauer von 15-20 min entwickelt.⁶⁷

Funktionalität wird anhand von Symptomen in 7 Stufen gemessen,

Inklusion von Items zu Partizipation als Ergebnisdimension

Duncan et al.⁶⁸ untersuchten die Testgütekriterien der Weiterentwicklung der ursprünglichen Stroke Impact Scale. Die SIS 2.0 besteht aus 64 Items in 8 Kategorien. 4 Kategorien betreffen Körper/Strukturen (Stärke, Handfunktion, Mobilität, Aktivitäten bzw. instrumentelle Aktivitäten des täglichen Lebens), die anderen vier sind Emotion, Kommunikation, Erinnerung und soziale Partizipation. Zur Identifikation jener Items, die die psychometri-

gute Reliabilität

sche Testgüte der SIS reduzieren, analysierten die Autoren einen RCT sekundär anhand der Rasch Analysemethode. (Dabei handelt es sich um ein vom dänischen Statistiker Rasch entwickelten Modell der probabilistischen Testtheorie.) Die Ergebnisse untermauern die Konstrukt-Validität des Instruments und führten zur Reduktion der Skala auf 16 Items. Die SIS 3.0 beinhaltet auch ein Item, bei dem der Heilungsfortschritt von Patient/innen selbst auf einer visuellen Analogskala einzuschätzen ist.⁶⁹

Die Stroke Impact Scale verfügt über ausgezeichnete Reliabilitätswerte, die von Duncan et al.⁷⁰ für die ersten 6 Monate nach Schlaganfällen an 91 Patient/innen getestet wurden und sich in einer anderen Studie bestätigten.⁷¹ Die einzige hinsichtlich ihrer Testgüte schwächelnde Kategorie ist „Emotion“ (Interclass Correlation Coefficient; $r=0,57$).

**systematische
Unterschätzung des
Gesundheitszustande,
aber hohe
Übereinstimmung mit
körperlicher
Funktionalität**

Die Testgütekriterien der SIS 2.0 sowie der SIS-16 wurden an einer australischen Studienpopulation erfolgreich getestet.⁷² Duncan et al.⁷³ untersuchten den Unterschied der Bewertung der Patient/innen auf der NIHSS-16 durch Personen, die die Patient/innen seit mindestens einem Jahr kennen und mindestens 1-mal wöchentlich kontaktieren ($N=377$). Ihre Bewertungen unterschätzten den Gesundheitszustand der Betroffenen systematisch, ein Bias, der mit zunehmendem Schweregrad steigt, mit Effektstärken von -0,1 bis 0,4. Der Zusammenhang (Interclass Correlation Coefficient) beträgt $r=0,5-0,83$, mit hoher Übereinstimmung für Bereiche der körperlichen Funktionalität.

**im Vergleich zu FIM &
SF-36 geringere Floor-/
Ceiling-Effekte**

In einer anderen Studie veröffentlichten die Autoren die Ergebnisse einer postalischen Fragebogenerhebung an einer Gruppe von Patient/innen, die seit mehr als 6 Monaten nach der Entlassung zu Hause lebten.⁷⁴ Der Rücklauf betrug 63 %, wovon die Hälfte der Fragebögen von Dritten ausgefüllt wurde und etwa 7 % der Werte fehlten. Die fehlenden Werte traten systematisch bei Patient/innen mit höherem Schweregrad auf. Abgesehen vom Information-Bias zeigten die Resultate geringere Floor- und Ceiling-Effekte als sie beim Functional Independence Measure/FIM und dem SF-36 auftraten.

**gute Reliabilität &
Validität**

Die Vor- und Nachteile der SIS, wenn sie telefonisch administriert wird, beschreiben Kwon et al.⁷⁵ Sie befragten Patient/innen von 9 unterschiedlichen Krankenanstalten 4 Monate nach dem Schlaganfall ebenfalls anhand der SIS, dem FIM und der SF-36. Die Autoren attestierten allen Instrumenten, keine Floor- oder Ceiling-Effekte aufzuweisen und unterstrichen das bessere Abschneiden des SIS als diagnosespezifischem Instrument bei Reliabilitäts- und Validitätskriterien.

**auch in deutscher
Version verfügbar**

Die SIS wurde von Petersen et al. in deutscher Sprache anhand der Antworten von 137 Personen psychometrisch getestet und validiert.⁷⁶

4.2.3 Barthel Index (BI)

**Beurteilung der
Fähigkeit für
Aktivitäten
10 Items**

Der Barthel Index (BI) wird seit 1955 verwendet.⁷⁷ Er wurde ursprünglich entwickelt, um die Fähigkeit von Patient/innen mit neuro-muskulären oder muskulo-skeletalen Störungen einfache Aktivitäten zu bewältigen, zu beurteilen. Der BI besteht aus 10 Items in 3 Kategorien (Mobilität, Körperpflege, Aktivitäten des täglichen Lebens), die innerhalb von 2-5 min von Patient/innen selbst bzw. innerhalb von etwa 20 min vom Behandlungspersonal bewertet werden können. Insgesamt können 100 Punkte erreicht werden – je höher der Score, desto größer die funktionale Unabhängigkeit.⁷⁸

Der BI erzielt unabhängig davon, ob das Behandlungspersonal für seine Anwendung trainiert wird, ähnlich reliable Ergebnisse.⁷⁹ Cohen et al. (2000) berichten in ihrem Review, dass Intrarater- ($r=0,71-0,99$) und Test-Retest-Reliabilität ($r=0,87$) ausgezeichnet hoch sind. Im Vergleich mit der Stroke Severity Scale und der Rankin Scale erzielte der BI bei D´Olhaberraigue et al. die höchste Interrater Reliabilität ($r=0,88$).⁸⁰ Auch die interne Konsistenz der Skala ist belegt.⁸¹ Die guten Reliabilitätswerte treten auch unabhängig von der Administrationsform auf. Bei Yeo et al.⁸² bestanden hohe Korrelationen zwischen telefonischen Interviews, postalisch versendeten Fragebögen und Face-to-Face Interviews zu Hause. Lediglich beim Erfassen der ADL durch Dritte ist die Evidenzlage leicht widersprüchlich.⁸³

Green et al. verglichen die Test-Retest Reliabilität des Barthel Index/BI, des Rivermead Mobility Index/RMI, der Nottingham Extended Activities of Daily Living Scale und des Frenchay Activity Index/FAI an Schlaganfallpatient/innen. Die Reliabilitätswerte von BI und RMI sind dabei besser als bei jenen Instrumenten, die vorwiegend die ICF-Dimension Partizipation messen (extended Activities of daily Living).⁸⁴

Der BI wurde häufig als „Gold Standard“ bei der Einschätzung der Validität anderer Ergebnisinstrumente bei Schlaganfall eingesetzt⁸⁵ und korreliert mit kurz- und langfristigem Outcome sowie den Hilfebedürfnissen und der Pflegeabhängigkeit zu Hause.^{86 87} Duncan et al.⁸⁸ testeten den BI in Kombination mit einschlägigen Symptomen als besten Schätzer des Gesundheitsstatus nach Schlaganfällen leichten Schweregrades.

Trotz der guten Reliabilität und Validität, guter Vorhersagekraft und einfacher Administrierbarkeit wurde der BI als Ergebnisinstrument der Schlaganfall-Rehabilitation wegen einer mangelnden Änderungssensitivität kritisiert. Wood-Dauphinee et al. errechneten jedoch eine Effektstärke von 0,42 fünf Wochen nach der akutstationären Aufnahme, und der BI schnitt in dieser Studie im Vergleich mit anderen Instrumenten am effektivsten ab.⁸⁹ In einer anderen Studie der Autorin erreichte der BI eine Effektstärke von 0,71 innerhalb von 6 Wochen und 0,25 zwischen 6-12 Wochen nach dem Schlaganfall, war aber der Balance-Scale leicht unterlegen.⁹⁰ Auch in der Studie von Stone et al.⁹¹ schnitt der BI akzeptabel ab, wenngleich die Änderungssensitivität lediglich anhand von p-Werten errechnet wurde, obwohl belastbarere Analysemethoden zur Analyse der Änderungssensitivität gefragt sind. Solche ROC/AUC-Analysen wurden in den Studien von Hocking et al.⁹² und Dromerick et al.⁹³ durchgeführt und erreichten Werte zwischen 0,66-0,81 für Veränderungen. In zwei weiteren Studien wurden moderate Effektstärken berechnet und erwartete Veränderungen konnten mit dem BI nachgewiesen werden^{94 95}, und das ähnlich gut wie mit dem Functional Independence Measure (FIM). Die Kritik an der Änderungssensitivität des BI erscheint insofern als fragwürdig, weil für viele andere Outcome-Instrumente der Rehabilitation nach Schlaganfall weit weniger Evidenz zur Beurteilung der Sensitivität vorliegt. Eine ernst zunehmende Schwäche zeigt der BI allerdings hinsichtlich auftretender Floor- und Ceiling-Effekte. Hobart & Thompson⁹⁶ berichten von eher geringen Ceiling-Effekten (1-5%), wenn jene fünf Items des 10 Item-BI exkludiert werden, die weit höhere Floor- und Ceiling-Effekte aufweisen. Ihre Population bestand aus Patient/innen einer Rehabilitationsklinik, die 5-Items-Kurzversion zuletzt aus „Transfer“, „Baden“, „Toilette“, „Stiegen“ und „Mobilität“. Die 5-Items-Kurzversion erreicht zwar hohe Reliabilität und Validität, man verliert jedoch die erstrebte Informationsdichte, und bei Hsueh et al.⁹⁷ tritt ein erheblicher Floor-Effekt auf. Bei Duncan et al.⁹⁸ beträgt der Ceiling-Effekt 66 %

**hohe Inter- &
Intrarater Reliabilität**

**schneidet auch bei
Vergleichen sehr gut ab**

**Barthel Index gilt als
„gold standard“**

**gute Vorhersagekraft
nach leichten
Schlaganfällen**

**Kritik an Mangel an
Änderungssensitivität**

**Diskussion um
Änderungssensitivität
ist aber fragwürdig, weil
Barthel Index gut
untersucht ist**

**BI zeigt Floor- &
Ceiling-Effekte**

bei allen Patient/innen mit leichten und mittelschweren Schlaganfällen, d.h. diese erreichten auf dem BI bereits den Höchstscore von 100. Eine andere Studie liegt mit einem Ceiling Effekt von 20-27 % bei Rehabilitand/innen ebenfalls über den akzeptablen Werten.⁹⁹ Hier schneidet das FIM wesentlich besser ab.

**leicht administrierbar &
gute Diskriminierung
zwischen selbständigen
& pflegebedürftigen
Patient/innen**

Insgesamt ist der Einsatz des BI eher aus rein klinischen Gesichtspunkten attraktiv, weil die Skala leicht einsetzbar ist und zwischen Patient/innen, die pflegerische Hilfe benötigen und selbständigen Patient/innen scharf trennt.¹⁰⁰ Für die Forschung und Qualitätsentwicklung in der Rehabilitation von Schlaganfallpatient/innen bestehen jedoch signifikante Schwächen aufgrund von Floor- und Ceiling-Effekten

4.2.4 Functional Independence Measure (FIM)

**zur Messung von
Behandlungs-&
Pflegeaufwand**

**USA: Basis für
Refundierung**

Der FIM wurde als Teil der Kritik am BI entwickelt. Das Instrument ist im Vergleich zu anderen nicht primär darauf ausgerichtet, Unabhängigkeit und Abhängigkeit zu messen, sondern die körperlichen und kognitiven Fähigkeiten der Patient/innen in Bezug zum Behandlungs- und Pflegeaufwand zu setzen. In diesem Sinne wird es in den U.S. heute als Basis der Refundierung in der Neurorehabilitation systematisch eingesetzt.¹⁰¹

**18 Items in 6 Bereichen
der Funktionalität**

Der FIM besteht aus 18 Items, welche 6 Bereiche der Funktionalität (Selbstversorgung (1) im Alltag, Kontinenz (2), Mobilität (3), Transfer (4), Kommunikation (5), soziale Kognition (6)) messen. Diese sechs Bereiche werden in zwei grundsätzlichen Kategorien zusammengefasst: körperliche Funktionalität (13 Items; angelehnt an den Barthel Index; auch „Motor-FIM“) und kognitive Funktionalität (5 Items; auch „Cognitive-FIM“). Alle Items sind auf einer 7er-Likert-Skala ordinalskaliert, wobei „1“ totale Hilfsbedürftigkeit und „7“ totale Unabhängigkeit beschreibt. So entsteht ein Summenwert von 18-126, wobei Linnacre et al.¹⁰² darauf hinweisen, dass die Verwertung der Subskalensummenwerte (motor-FIM, cognitive-FIM) in vielen Fällen nützlicher ist als anhand der Gesamtsumme zu werten.

**generisches Instrument
über alle Neuro-Rehab
Diagnosen,**

Der FIM kann generisch für alle Diagnosegruppen in der Rehabilitation eingesetzt werden. Die Administration des FIM erfordert Training und Zertifizierung. Die Reliabilität des FIM ist direkt von der Fähigkeit der Beobachter/innen, mit dem FIM umzugehen, abhängig, weshalb die Notwendigkeit von Training nicht unterschätzt werden sollte. Üblicherweise werden die Items vom Behandlungspersonal auf der Basis von Beobachtung fremdeingeschätzt, wobei optimale Einschätzungen auf der Grundlage gemeinsamer Beobachtung in interdisziplinären Teams erfolgen. Das Ausfüllen des FIM benötigt in etwa 30 min. Andere Administrationsarten führen zu einer Reduktion der Testgütekriterien, insbesondere in der Kategorie Cognitive-FIM.

bedarf aber Training

**gute Reliabilität &
Validität**

Dodds et al.¹⁰³ wiesen an einer Studienpopulation von mehr als 11.000 Patient/innen exzellente Reliabilitätswerte und gute Validität des Instruments nach, wobei auch Schlaganfall-Patient/innen involviert waren.

**Vorhersagekraft für
Pflege- und
Betreuungsaufwand ist
gut**

Der FIM sagt mit einiger Genauigkeit den Verbleib der Patient/innen nach der Entlassung aus dem akutstationären Bereich voraus und in Kombination mit der Aufenthaltsdauer auch die Verbesserung der Funktionalität.¹⁰⁴ Granger et al.¹⁰⁵ prognostizierten anhand der FIM den Pflege- und Betreuungsaufwand zu Hause sowie anhand des FIM und der Erhebung von Symptomen bei Entlassung sowie der Sehkraft auch die Lebenszufriedenheit. Sie

errechneten, dass eine Änderung auf dem globalen FIM-Score von einem Punkt innerhalb der FIM-Score-Bandbreite von 61 und 126 einer Veränderung von 2,19 Minuten Pflegeaufwand bedeuten.

Brock et al. veröffentlichten die Vorteile in der Unterscheidung der Schweregrade von Schlaganfall-Patient/innen im Vergleich zu anderen Instrumenten.¹⁰⁶ Der FIM unterscheidet auch zwischen Schweregraden bei Rückenmarkverletzten, Komorbidität, rechts- und linksseitig beeinträchtigenden Schlaganfällen und sowohl bei Aufnahme als auch bei Entlassung aus stationären Rehabilitationseinrichtungen¹⁰⁷ sowie zwischen Patient/innen mit und ohne Aphasie.¹⁰⁸

Die Änderungssensitivität des FIM bei seinem Einsatz an Schlaganfall-Patient/innen wird in der Literatur wie folgt beschrieben: Dodds et al.¹⁰⁹ finden Veränderungen bei Patient/innen wie erwartet. Für Motor-FIM berichten Brock et al.¹¹⁰ 16 % Ceiling-Effekt, van der Putten et al.¹¹¹ finden keine signifikanten Verzerrungen des FIM bei Schlaganfall-Patient/innen. Wallace et al.¹¹² errechneten eine Effektstärke von 0,31 (0,46 bei Patient/innen mit bekannten Veränderungen, entspricht einer mittleren Effektstärke) und AUC/ROC-Kurve=0,675. Dies entspricht einer Wahrscheinlichkeit von 67,5 %, dass das FIM eine zufällig gewählten Patient/in, bei der eine Verbesserung des Gesundheitszustandes stattgefunden hat, auch als solche erkennt. Dromerick et al.¹¹³ finden ein Standardized Response Mean = 2,18 zwischen Aufnahme und Entlassung – der FIM misst die vorhergesagten Veränderungen bei 91 von 95 Patient/innen, das sind 18 mehr als im Vergleich zum Barthel Index. Dieser Vorteil rührt aus der größeren Bandbreite von Antwortmöglichkeiten und dem dichteren Informationsniveau des FIM. Die Evidenz zur Änderungssensitivität des FIM ist allerdings leicht widersprüchlich, denn in der vergleichenden Evaluation der Änderungssensitivität von FIM, Motor-FIM und BI finden van der Putten et al. keine signifikanten Unterschiede.¹¹⁴

Schweregrad-Unterscheidung bei verschiedenen Diagnosegruppen möglich

Änderungssensitivität; etwas widersprüchliche Aussagen,

FIM: breitere Messung des Gesundheitszustands als BI, beide Instrumente Floor- & Ceilingeffekte

BI leichter zu administrieren,

Einsatz hängt von Akzeptanz ab

Table 2 BI and FIM scores on admission: sample range, mean, floor, and ceiling effect in two disease groups

Scale range	n	Sample range	Admission score mean (SD)	Floor effect n (%)	Ceiling effect n (%)
<i>BI (0-20)</i>					
MS	201	0-20	12.0 (5.7)	3 (1.5)	11 (5.5)
Stroke	82	0-20	11.4 (5.5)	1 (1.2)	7 (8.5)
<i>FIM total (18-126)</i>					
MS	201	24-122	89.4 (23.0)	0(0)	0 (0)
Stroke	82	21-123	82.5 (26.8)	0 (0)	0 (0)
<i>FIM motor (13-91)</i>					
MS	201	13-88	59.1 (20.4)	4 (2.0)	0 (0)
Stroke	82	13-91	56.1 (20.9)	1 (1.2)	1 (1.2)
<i>FIM cognitive (5-35)</i>					
MS	201	11-35	30.3 (5.1)	0 (0)	36 (17.9)
Stroke	82	5-35	26.4 (7.9)	1 (1.2)	11 (13.4)

MS=multiple sclerosis; floor (ceiling) effects are the percentage of patients scoring the minimum (maximum) possible scores.

Table 3 Comparison of BI and FIM change scores, p values, and effect sizes

Disease groups	Change score mean (SD)	p Value	Effect size
<i>MS patients</i>			
BI	2.1 (2.4)	< 0.0001	0.37
FIM total	6.9 (8.3)	< 0.0001	0.30
FIM motor	6.9 (7.2)	< 0.0001	0.34
FIM cognitive	0.1 (2.9)	0.961 (NS)	0
<i>Stroke patients</i>			
BI	5.2 (4.4)	< 0.0001	0.95
FIM total	21.9 (19.0)	< 0.0001	0.82
FIM motor	19.1 (16.1)	< 0.0001	0.91
FIM cognitive	2.8 (4.8)	< 0.0001	0.61

Effect size=change score/standard deviation of admission score.

Abb. 4.2-2: Vergleich unerwünschter Informationseffekte von Barthel Index und Functional Independence Measure

Die Ergebnisse der an 201 Patient/innen mit Multipler Sklerose und 82 Schlaganfallpatient/innen durchgeführten Studie sind in Abb. 4.2-2 (Table 2) dargestellt. Der FIM-Summenwert weist bei keiner/m Patient/in Floor- oder Ceiling-Effekte auf, nennenswerte Probleme treten in der Subskala Cognitive-FIM sowie beim BI und hier eher bei Patient/innen mit multipler Sklerose auf. Im unteren Teil der Abbildung 4.2-2 (Table 3) werden einander die Effektstärken gegenübergestellt, die höheren Werte deuten darauf hin, dass das FIM insbesondere zur Veränderungsmessung bei Schlaganfallpatient/innen taugt. Die größte Effektstärke erzielt jedoch der BI.

Dies würde bedeuten, dass das FIM keine echten Vorteile gegenüber dem BI aufweisen kann, sieht man von der genaueren und breiteren Messung des Gesundheitszustands auf dem FIM ab. Eine Schwäche, die beide Instrumente aufweisen, sind Floor- und Ceiling-Effekte, die auf den ersten Blick zwar nicht übermäßig bedeutsam erscheinen. Die Instrumente wurden jedoch nur in Bezug auf ihre Fähigkeiten, positive Veränderungen des Gesundheitszustands zu messen, also eindimensional evaluiert, über ihre Fähigkeit zur Messung von Verschlechterungen ist nichts bekannt. Im Vergleich zum FIM ist der BI wesentlich einfacher und flexibler zu administrieren. Da die beiden Instrumente sehr ähnliche psychometrische Qualität besitzen, sollte der Einsatz in der Praxis von der Akzeptanz bei Personal und Patient/innen abhängig gemacht werden.

Schepers et al. verglichen die Änderungssensitivität des BI, FIM, FAI und SA-SIP 30 in einer Studie mit 163 Patient/innen, die zum ersten Mal einen Schlaganfall erlitten.¹¹⁵ Die Effektstärken waren in der subakuten Phase bei allen Instrumenten ähnlich, Cognitive-FIM zeigte einen beträchtlichen Ceiling-Effekt und hatte nur durchschnittliche Effektstärke. In der chronischen Phase eruierten FAI und SA-SIP-30 die Veränderungen des Gesundheitszustands am ehesten, weshalb die Autoren diese Instrumente für die Rehabilitation empfehlen.

Effektstärken und Veränderungssensitivität bei subakuten & chronischen Patient/innen unterschiedlich

4.2.5 Frenchay Activities Index (FAI)

Der Frenchay Activities Index/FAI misst diagnosespezifisch instrumentelle Aktivitäten des täglichen Lebens, Dabei geht er über den Rahmen anderer Instrumente, die sich schwerpunktmäßig auf Pflege(un)abhängigkeit und Mobilität beziehen, hinaus, indem er den Lebensstil von Patient/innen miteinbezieht (Hausarbeiten, Freizeit- und Berufsaktivitäten und Aktivität im Freien).¹¹⁶ Wade et al.¹¹⁷ evaluierten die Skala im Hinblick auf ihre Testgüte an 976 Patient/innen und fanden heraus, dass niedrigere Summenwerte (15-60 bzw. 0-45 aus 15 Items) mit Depression, niedrigem IQ und – 1 Jahr nach dem Schlaganfall – mit dem Geschlecht korrelieren (Frauen).

misst Aktivitäten des täglichen Lebens/ ADL diagnosespezifisch

Der FAI kann in einem Interview innerhalb von 5 min ausgefüllt werden, aber auch postalische Versendung von Fragebögen funktioniert bei sehr guter Reliabilität. Von Dritten (Angehörigen) sollte der FAI eher nicht komplettiert werden, da sie die Fähigkeiten der Patient/innen systematisch unterschätzen.¹¹⁸

Reliabilität und Validität des FAI sind ausgezeichnet und auch relativ gut evaluiert. Die Test-Retest-Reliabilität beträgt in der Studie $r=0,8$, die Inter-Rater Reliabilität bei Piercy et al.¹¹⁹ zwischen 0,27 für soziale Aktivitäten und 0,8 für Aktivitäten im Haushalt, für die Summenwerte $r=0,93$. Der FAI korreliert mit FIM und BI ($r=0,66-0,7$) sowie mit dem Sickness Impact Profile (SIP) ($r=0,7$), was belegt, dass das Instrument misst, was es messen soll.

ausgezeichnete Reliabilität und Validität,

Schuling et al.¹²⁰ finden weder Floor- noch Ceiling-Effekte, und auch Wade et al. belegen, dass die Summenwerte gut verteilt sind und Veränderungen des Gesundheitszustands nach 6 Monaten in der erwünschten Richtung messen. Das diagnosespezifische Instrument zur Bestimmung der ADL bei Schlaganfallpatient/innen schneidet sehr gut ab, vor allem was die Änderungssensitivität betrifft, ergibt sich ein Vorteil gegenüber dem BI oder dem FIM.

hohe Korrelation mit anderen vergleichbaren Instrumenten

gute Änderungssensitivität

4.2.6 Mini-Mental State Examination (MMSE)

Mini-Mental State Examination/MMSE wurde ursprünglich als Testinstrument für kognitive Einschränkungen bei Demenzen entwickelt,¹²¹ wird mittlerweile aber auch bei anderen Diagnosegruppen eingesetzt. Die Skala besteht aus 11 einfach zu verstehenden Fragen und Aufgaben, die in 7 Kategorien differenziert werden: zeitliche (1) und örtliche (2) Orientierung, Merken und Widergeben von drei Wörtern (3), Aufmerksamkeit (4), Fähigkeit, einfache Rechenaufgaben zu lösen (5), Sprache (6) und visuelle Fähigkeiten (7). Trainiertes Personal kann die Skala in Zusammenarbeit mit Patient/innen in etwa 10 min. ausfüllen.

misst kognitive Befähigung; 11 Fragen in 7 Kategorien

Ein Ergebnis von 23 oder weniger Punkten (von insg. 30) wird allgemein als Cut-off für kognitive Einschränkungen akzeptiert.¹²² Ein Score unter 17 gilt als schwere Beeinträchtigung.¹²³ Teng & Chui¹²⁴ entwickelten eine erweiterte Version der MMSE, die “modified mini mental state examination” (3MS), die bei anderer Skalierung ca. 15 min pro Patient/in erfordert.

Reliabilität und Validität der Skala sind hinlänglich erwiesen.¹²⁵

**gute Reliabilität &
Validität
schwache
Vorhersagekraft**

Ozdemir¹²⁶ macht auf die schwache Vorhersagekraft der Skala aufmerksam, in dem er sie während Krankenhausaufenthalt (Aufnahme-Entlassung) mit den Werten des FIM verglich ($r=0,31$). Die Evidenzlage hinsichtlich der Frage, ob Alter, Bildungsgrad und sozio-ökonomischer Status die Ergebnisse auf der MMSE systematisch beeinträchtigen, ist ungeklärt. Lorentz et al.¹²⁷ befürchten, dass Adjustierung nach diesen Variablen den Nutzen der Skala reduzieren.

**geringe
Änderungssensitivität**

Neben geringer Änderungssensitivität und diskriminativer Validität ist für die Anwendung der MMSE bei Schlaganfall auch bedeutsam, dass die Inhaltsvalidität schwach ist. Items, die Sprache und Visualität betreffen, können nicht bei allen Schlaganfallpatient/innen komplettiert werden.¹²⁸ Ob diese Schwächen teilweise wettgemacht werden können, wenn das Set etwa um den Uhrzeiger-Test erweitert wird, ohne dass die Vorteile der Skala (einfach, schnell) darunter leiden¹³², wird von Kliniker/innen bezweifelt, weil die Defizite von Schlaganfallpatient/innen die Aussagekraft einschränken.¹²⁹

4.2.7 Berg Balance Scale (BBS)

**misst die Balance-
Fähigkeit
14 Items**

Die Berg Balance Scale/ BBS misst die Balance-Fähigkeit älterer Erwachsener.¹³⁰ Die Skala besteht aus 14 Items. Die Patient/innen werden aufgefordert, in körperlichen Positionen zu verharren und Übungen verschiedenen Schwierigkeitsgrades durchzuführen, die üblicherweise auch im täglichen Leben durchgeführt werden. Zur Administration benötigt man ein Lineal, eine Stoppuhr, einen Sessel und 10-15 min Zeit. Die Bewertungen werden durch Beobachtung vorgenommen. Jedes Item wird zwischen 0 und 4 bewertet. Erreichen die Patient/innen weniger als 45 von 56 möglichen Punkten, gilt ihre Balance als beeinträchtigt.¹³¹ Ein Wert unter 20 Punkten ist im klinischen Alltag als „rollstuhlpflichtig“ definiert.

In ihrem systematischen Review bewerten Blum & Korner-Bitensky die psychometrischen Testgütekriterien der BBS auf der Basis von 21 Studien. Die Reliabilitäts- und Validitätswerte der BBS sind exzellent (interne Konsistenz, Inter- und Intrarater Reliabilität, Test-Retest-Reliabilität). Korrelation bestehen mit dem Barthel Index, der Postural Assessment Scale for Stroke Patients, der Balance-Subskala des Fugl-Meyer-Assessments und dem FIM. Die BBS scheint in der Lage, die Aufenthaltsdauer, die nachfolgende Behandlungsorganisation sowie die körperliche Funktionalität nach 90 und 180 Tagen vorherzusagen, nicht aber die Wahrscheinlichkeit für Stürze.¹³²

gute Reliabilität

Einer anderen Studie zufolge ist die Test-Retest Reliabilität bei Überlebenden von Schlaganfällen sehr hoch ($r=0,99$), Interrater Reliabilität ($r=0,98$) und interne Konsistenz sind ebenfalls sehr hoch, während die konkurrente Validität zu anderen Ergebnisinstrumenten mittelmäßig ausgeprägt ist.¹³³ Auch andere Validitätswerte sind moderat. Hier ist die BBS mit einer Sensitivität von 0,56 und Spezifität von 0,96 in der Lage, Patient/innen, die nach der Entlassung stürzen werden, zu identifizieren.

Bezüglich der Änderungssensitivität schneidet die BBS zwar gut ab (ES=0,8 zwischen 14 und 30 Tagen und beträgt noch 90 bis 100 Tage nach dem Schlaganfall ES=0,4).^{134 135} Im Vergleich zu anderen Instrumenten, die die Balance messen, treten jedoch teilweise erhebliche Floor- und Ceiling-Effekte auf (vgl. Abb. 4.2-3 unten).

Die BBS ist bei stark beeinträchtigten Schlaganfall-Patient/innen nur beschränkt einsetzbar, da nur eines der Items im Sitzen durchgeführt werden kann. Aktive Patient/innen empfinden die Übungen hingegen als zu einfach.

**kurzfristige
Änderungssensitivität,
Floor- & Ceiling Effekte**

**nur bedingt für schwere
Schlaganfall-
Patient/innen geeignet**

4.2.8 Fugl-Meyer Assessment of Motor Recovery after Stroke (FMA)

Das Fugl-Meyer Assessment (FMA) ist ein diagnosespezifischer Index, um motorische Funktionalität, Balance und sensorische Empfindungen bei halbseitig gelähmten Schlaganfall-Patient/innen zu testen und fügt mehrere zuvor bestehende Skalen modular zusammen. Die Skala besteht aus 5 Kategorien, die schwerpunktmäßig der ICF-Klasse 1 (body/function) zugeordnet werden können. Die Items werden auf einer 3er-Ordinalskala bewertet, sodass insgesamt 226 Punkte möglich sind. Items, die motorische Funktionalität werten, sind dabei im Vergleich zu den anderen Kategorien stark übergewichtet. Eine Klassifikationssystematik zur Schweregradifferenzierung besteht.¹³⁶

**misst motorische
Funktionalität, Balance
& sensorische
Empfindungen in 5
Kategorien**

Gladstone & Dannels¹³⁷ berichten in ihrem Review, dass es durchaus üblich ist, die Skalen voneinander getrennt einzusetzen, insgesamt dauert die Einschätzung auf dem FMA 30-45 min. und wird mittels Patient/innen-Beobachtung von Physiotherapeut/innen durchgeführt.

Mao et al.¹⁴⁰ verglichen die Sub-Skala zur Balance (FM-B) mit zwei anderen Balanceskalen, die in der Schlaganfall-Rehabilitation eingesetzt werden. Während Konstrukt- und Konvergenzvalidität der Skala sehr gut sind, schneidet die Reliabilität durchschnittlich ab, was an mangelnder Schlaganfall-Spezifität der Items liegen mag. Bei einem Vergleich der Änderungssensitivität aller drei verglichenen Skalen schneidet die FM-B jedoch relativ gut ab. Die bestehenden Ceiling-Effekte, sind aus Sicht der Messtheorie in ihrem Ausmaß vertretbar. Werden bei bestimmten Subgruppen jedoch Verschlechterungen des Gesundheitszustands erwartet, schneidet die Subscala zur Balance mit einem Floor-Effekt von 36 % als valides Messinstrument jedoch aus (vgl. Abb. 4.2-3).

**nur durchschnittliche
Reliabilität,
Änderungssensitivität
gut**

TABLE 2. Distributions of the 3 Balance Measures at Different Stages of Stroke Recovery

	PASS, n (%)		FM-B, n (%)		BBS, n (%)	
	Floor Effect	Ceiling Effect	Floor Effect	Ceiling Effect	Floor Effect	Ceiling Effect
14 DAS (n=123)	10 (8.1)	4 (3.3)	36 (29.3)	4 (3.3)	43 (35)	6 (4.9)
30 DAS (n=110)	5 (4.5)	11 (10)	16 (14.5)	11 (10)	19 (17.3)	13 (11.8)
90 DAS (n=93)	2 (2.2)	13 (14)	6 (6.5)	13 (14)	6 (6.5)	20 (21.5)
180 DAS (n=80)	3 (3.8)	14 (17.5)	5 (5)	14 (17.5)	4 (5)	23 (28.8)

Table 2. Distributions of the 3 Balance Measures at Different Stages of Stroke Recovery

From: Mao: Stroke, Volume 33(4).April 2002.1022-1027

Abb. 4.2-3: Unerwünschter Informationseffekte bei verschiedenen Balance-Messungen

Die Abb. 4.2-3 (Table 2) vergleicht drei Messinstrumente (PASS, Fugl-Meyer-Subskala zur Balance und Berg Balance Scale) hinsichtlich Floor- und Ceiling Effekten in verschiedenen Rehabilitationsstufen (DAS). Abhängig von dieser Einschätzung des Gesundheitszustandes der Patient/innen erzielt die Subskala zur Balance auf dem FMA bessere Werte als die Berg Balance Scale, alle Werte liegen jedoch in einem akzeptablen Bereich.

**sehr zeitaufwändig,
beschränkt auf
körperliche
Funktionalität**

Wenn das FMA in voller Länge eingesetzt wird, ist das Assessment nur mit hohem Zeitaufwand zu administrieren, was seine Einsetzbarkeit im klinischen Behandlungsalltag beschränkt. Dass die Skala auf körperliche Funktionalität eingrenzt und nur eine ICF-Kategorie einschließt, führt dazu, dass funktionelle und partizipative Veränderungen der Patient/innen nur unzureichend abgebildet werden können.

4.2.9 Medical Outcomes Study Short Form 36 (SF-36)

**generisches Instrument
zum Gesundheitsstatus
36 Items in 8 Kategorien**

Der SF-36 ist ein generisches Instrument und misst den Gesundheitsstatus.¹³⁸ Es besteht aus 36 Items, die sich aus den ursprünglichen 245 Items der Skala generieren. Die 8 Kategorien sind: körperliche Funktionalität und Einschränkungen, Schmerzen, soziale Rollenfunktion, generelle psychische Gesundheit und Einschränkungen, Vitalität und genereller Gesundheitszustand. Zwei Items erfragen die Veränderung des Gesundheitszustandes innerhalb des letzten Jahres. Mit Ausnahme dieser zwei Fragen beziehen sich die Items auf die 4 Wochen vor der Befragung, darüber hinaus gibt es eine Version, die sich nur auf die letzte Woche bezieht. Für jede der 8 Kategorien kann ein Summenwert von 100 erzielt werden und es gibt verschiedenste Referenzwerte, auch für Schlaganfall-Patient/innen.¹³⁹

**Reliabilität sinkt bei
Ausfüllen durch Dritte**

Der SF-36 kann als Selbstausfüller-Fragebogen oder als persönliches oder telefonisches Interview administriert werden. Das Ausfüllen dauert weniger als 10 min.¹⁴⁰ Von einer Erhebung durch Dritte ist abzuraten, die Reliabilität sinkt teilweise dramatisch.¹⁴¹

gute Reliabilität

Dorman et al. verglichen die Reliabilität des SF-36 mit jener des EuroQuoL an 2253 Schlaganfall-Patient/innen und attestieren ähnlich gute Werte, mit

Ausnahme der Kategorie „psychische Gesundheit“.¹⁴² Die Autoren merken an, dass beide Instrumente nicht für serielle Beobachtungen eingesetzt werden sollten, wenn nicht sehr große Veränderungen des Gesundheitszustandes erwartet werden. Anderson et al. validierten die Skala bei 90 tendenziell älteren Schlaganfall-Patient/innen (Durchschnittsalter 72 Jahre) im Vergleich zu anderen Instrumenten, mitunter dem Barthel Index. Der SF-36 schnitt dabei mittelmäßig ab, insbesondere die Skala zu sozialen Rollenfunktionen bereitete Probleme.¹⁴³ Bei älteren Patient/innen sind auch relativ hohe Raten an fehlenden Daten zu erwarten, wenn der SF-36 postalisch versendet wird.^{144 145 146} Manche Autoren sind der Meinung, dass dies mit der Verständlichkeit der Skala zu tun hat.^{147 148}

Bzüglich Änderungssensitivität weist der SF-36 in vielen Studien hohe Ceiling-, dafür kaum Floor-Effekte auf.¹⁴⁹ Bei Studien mit Schlaganfall-Patient/innen hingegen treten die Ceiling-Effekte geringer zu Tage als vergleichsweise beim Barthel Index.¹⁵⁰

Der SF-36 ist einfach und flexibel administrierbar und hat sich in vielerlei Kontexten bewährt, weshalb auch ausreichende Referenzwerte vorliegen. Aber auch die Kurzversion mit 12 Items wurde eingesetzt und hat sich als reliabel und valide erwiesen.^{151 152} Die Änderungssensitivität ist gering bis mittel ausgeprägt und steigt mit dem Schweregrad der Symptome tendenziell an.¹⁵³

**Änderungssensitivität:
hohe Ceiling-Effekte**

4.2.10 Stroke Specific Quality of Life Scale (SSQOL)

Die Stroke Specific Quality of Life Scale/SSQOL misst die gesundheitsbezogene Lebensqualität von Schlaganfall-Überlebenden in einem Selbstausfüller-Fragebogen mit 49 Items in 12 Kategorien: Mobilität, Energie, Funktionalität der oberen Extremitäten, Arbeitsfähigkeit, Stimmung, Pflegeabhängigkeit, soziale und familiäre Rollenfunktionalität, Lebensziele, Sprache, Denken und Persönlichkeit. Jedes Item wird in einer 5er-Likert-Skala gemessen,¹⁵⁴ wobei höhere Summenwerte bessere Werte bedeuten. Die Summenwerte sind ungewichtet anwendbar.

Bei Williams et al.¹⁵⁵ die die Skala entwickelten, ist die Reliabilität des Instruments hoch und die Validität durchschnittlich, was auf die Subskalen über soziale Rollenfunktionen, Sprache, Denken und Lebensziele zurückzuführen ist. Auch die Effektstärken lassen zu wünschen übrig. Zudem berichten Czwchowsky & Hill¹⁵⁶ über Ceiling-Effekte, die 20 % übersteigen.

**misst Lebensqualität,
49 Items in 12
Kategorien**

**gute Reliabilität,
durchschnittliche
Validität, hohe Ceiling
Effekte**

**misst diagnose-
spezifisch
Gesundheitsstatus**

64 Items in 8 Kategorien

4.2.11 Die EuroQol Quality of Life Scale (EQ-5D)

Die EuroQol Quality of Life Scale/ EQ-5D ist ein generisches Instrument zur Erfassung und gesundheitsökonomischen Bewertung gesundheitsbezogener Lebensqualität. Es ist äußerst kurz und kann in verschiedenen Administrationsformen (Selbstausfüller, Interview) innerhalb von 2-5 min. erhoben werden.¹⁵⁷ Der erste Teil der Skala besteht aus 15 Items, jeweils 3 zu den Kategorien Mobilität, Selbsthilfe, ADL, Schmerzen und Depression, die 3 Antwortmöglichkeiten bieten (keine oder kaum Probleme, mittelmäßige bzw. extreme Probleme). Durch eine Transformation der Antworten der Gesamtbevölkerung entsteht so ein Lebensqualitätsindex, aus dem heraus einem Gesundheitszustand ein Wert zwischen 0-1 zugewiesen werden kann, welcher in gesundheitsökonomischen Evaluationen neuer Therapien Be-

misst Lebensqualität

15 Items in 3 Kategorien

**wird für gesundheits-
ökonomische Evaluation
herangezogen**

rücksichtigung findet. Für Österreich existieren keine Standardwerte.¹⁵⁸ Der zweite Teil der EQ-5D-Skala besteht aus einer visuellen Analogskala zur Selbsteinschätzung des Gesundheitszustands zwischen 0 und 100.

**Problem: fehlende
Werte für ältere
Menschen**

Manche Autoren empfehlen, den EQ-5D nicht zum Assessment individueller Patient/innen, sondern eher in Vergleichsstudien von Patient/innenkollektiven anzuwenden.¹⁵⁹ Ein anderes Problem stellen fehlende Werte von Älteren (10 %, Brazier et al.¹⁶⁰) dar. Coast et al.¹⁶¹ meinen, dass dies für 11 % der 65jährigen und 73 % der über 85jährigen notwendig sei.

Probleme bereiten auch Erhebungen, die nicht direkt an Patient/innen, sondern über Informationen Dritter durchgeführt werden. In einer weiteren Studie von Dorman et al.¹⁶² konnten 40 % der Patient/innen die Skala nicht ohne Hilfe des Pflegepersonals komplettieren, wobei sowohl Selbstausfüller als auch die assistierten Patient/innen zu ähnlich guter Reliabilität und Validität führten.

4.2.12 Nottingham Health Profile (NHP)

**misst soziale und
individuelle
Partizipation

2 Teile**

**Teil 1: 38 Items in 6
Kategorien zur
subjektiven Gesundheit**

Teil 2: Aktivitäten

Das Nottingham Health Profile/ NHP ist ein außergewöhnliches Instrument, um die subjektive soziale und individuelle Partizipation bei gesundheitlichen Beeinträchtigungen zu erfassen, das von Hunt und anderen zwischen 1980 und 1985 entwickelt wurde.¹⁶³ Es ist weder darauf ausgerichtet, gesundheitsbezogene Lebensqualität noch den Gesundheitszustand zu messen, sondern zielt direkt auf Items, die die Qualität der Beeinträchtigung selbst quantitativ erfassen. Das NHP besteht aus zwei Teilen: Teil 1 mit 38 Items in 6 Kategorien der subjektiven Gesundheit: Mobilität, Schmerz, Schlaf, soziale Isolation, emotionale Reaktion und Energielevel. Die Items messen Zustimmung zu Aussagen über jeweils relevante Probleme, die mit „Ja“ oder „Nein“ beantwortet werden sollen. Die Statements sind nach Schweregrad von Beeinträchtigungen gewichtet, so dass ein Index in jeder Kategorie von 100 entsteht (je höher der Score, desto geringer die subjektive Gesundheit).

Teil 2 besteht aus sieben Items über Aktivitäten, die vom Gesundheitszustand der/s Befragten beeinträchtigt werden können: Erwerbsarbeit, Haushaltsaktivität, soziales Leben, persönliche Beziehungen, Sexualleben, Hobbies und Freizeitaktivitäten, Urlaub. Die Skalierung erfolgt über eine dichotome „ja“-„nein“ Skala, je nachdem, ob sich die/der Befragte im jeweilige Bereich beeinträchtigt fühlt. Auch dieser Teil der Skala wurde sprachvalidiert.¹⁶⁴

**durchschnittliche
Reliabilität**

Das NHP wird als Selbstausfüller-Fragebogen oder als Interview durchgeführt und benötigt etwa 10 min, um vervollständigt zu werden. Die Administration wird von einem User-Manual¹⁶⁵ unterstützt, die Auswertung kann mit Referenzgruppen (Alter, Geschlecht, soziale Schicht) verglichen werden.¹⁶⁶

Die Test-Retest-Reliabilität rangiert in verschiedenen Studien mit Osteoarthritis- oder Herzinfarktpatient/innen zwischen 0,44 und 0,86 in verschiedenen Kategorien.¹⁶⁷ Visser et al.¹⁶⁸ führten eine der wenigen Arbeiten durch, die das NHP an Schlaganfall-Patient/innen anwendeten, lediglich 16 von insg. 53 Patient/innen (die anderen überlebten Herzinfarkt bzw. gehörten einer Kontrollgruppe an) litten an den Folgen eines Schlaganfalls. Die Test-Retest-Reliabilität betrug $r=0,65$ (Schmerz) bis $r=0,88$ (Teil 2 des NHP).

Die Skala erreicht hohe Testgüte bei anderen Reliabilitätskriterien, getestet an anderen als Schlaganfall-Patient/-innen.¹⁶⁹

Ebrahim et al., die den Einsatz der Skala bei Schlaganfall validierten, argumentieren, dass das NHP einfacher als andere Skalen, die für Schlaganfall-überlebende zu umfangreich seien, einzusetzen ist.¹⁷⁰ Die NHP hat sich bei den 463 Patient/-innen insbesondere dafür geeignet, Depression richtig einzuschätzen. Allerdings handelte es sich um eine Patient/-innengruppe, deren subjektive Gesundheit sich tendenziell verschlechterte (das Durchschnittsalter in der Kontrollgruppe betrug 68,8-71,8 Jahre). Zwischen NHP-Scores und Alter, Geschlecht bzw. Wohnstatus bestanden eher geringe Korrelationen.

Salter et al.¹⁷¹ bewerten die Validität des NHP durchschnittlich. Bezüglich Änderungssensitivität bestehen Hinweise auf erhebliche Ceiling-Effekte bei rheumatischen Patient/-innen und Rückenmarksverletzten, die auf Schlaganfallpatient/-innen nicht zutreffen müssen. Obwohl das NHP im Vergleich mit ähnlichen Instrumenten zur Erfassung der gesundheitsbezogenen Lebensqualität einfach und relativ flexibel einzusetzen ist (Datengenerierung bei Angehörigen abzurufen), dürfte die Anwendbarkeit bei Schlaganfallpatient/-innen mit niedrigem Schweregrad beschränkt sein.

Eine multizentrische Studie ging der Frage nach, ob Veränderungen der gesundheitsbezogenen Lebensqualität von Schlaganfallpatient/-innen objektiv sind oder nur auf Veränderungen der Bewertungsmaßstäbe der Patient/-innen zurückzuführen sind.¹⁷² Dabei wurden Schlaganfallpatient/-innen mit einer Kontrollgruppe verglichen, ohne entsprechende Hinweise auf einen "Response Shift" zu finden.

weniger umfangreich als andere Skalen

Validität durchschnittlich, Änderungssensitivität: erhebliche Ceiling-Effekte

wenig Erfahrung mit Schlaganfall

Tab. 4.2-4: Übersicht der Skalen und Scores bei Schlaganfall

STROKE

Instrument	Bez.	ICFclass	Admin	pop	Reliabilität	Validität	Änderungs- sensitivität	Quelle
National Institute of Health Stroke Scale NIHSS		1	O	n=43	+++	N.A.	N.A.	Berger 1999
		1	O	n=51	+++	++	N.A.	de Caneda 2006
				n=11	+	N.A.	N.A.	Josephson 2006
European Stroke Scale	ESS	1	O	n=43	+++	N.A.	N.A.	Berger 1999
Stroke Impact Scale 2.0	SIS	1,2,3	S	n=91	+++	++	+	Duncan 1999
			S	n=696	+++	+++	N.A.	Duncan 2003
			S	n=74	+++	++	N.A.	Edwards 2003
			S	n=203	N.A.	++	+	Duncan 2002
Mini Mental State Examination	MMSE	1	I	N.A.	++	+++	N.A.	Tombaugh 1992
			I	n=43	N.A.	N.A.	+	Ozdemir 2001
Modified MMSE	3MS	1	I	n=147	N.A.	+++	N.A.	Grace 1995
	FM-B	1	O	n=406	++	+++	++	Mao 2002
Barthel Index	BI	1,2	var.	n=25	+++	N.A.	N.A.	Collin, 1988
			var.	N.A.	+++	+++		Cohen, 2000
			O	N.A.	+++	+++		D'Olhaberriague 1996
			O	n=121	+++	++	++	Hsueh 2001
			O	n=844	N.A.	++	++	Hobart 2001
			O	n=697	N.A.	++		Duncan 1997
			O	n=142			+	Wood-D. 1990
			O	N.A.	N.A.	N.A.	++	Wood-D. 1996
			O	n=102	N.A.	N.A.	+	Stone 1995

		I	n=59	N.A.	N.A.	++	Dromerick 2003
		I	n=201	N.A.	N.A.	++	van der Putten 2001
Functional Independence Measure	FIM		n=11102	+++	++	+	Dodds 1993
		I	n=14799	N.A.	++	N.A.	Linacre 1994
		I	n=126	N.A.	+++	N.A.	Brock 2002
		I	n=93	N.A.	N.A.	++	Dromerick 2003
		I	n=201	N.A.	N.A.	++	van der Putten 2001
Frenchay Activities Index	(FAI)	I	n=976	+++	+++	+++	Wade 1985
		I	n=59	++	N.A.	N.A.	Piercy 2000
		I	n=185	+++	+++	N.A.	Schuling 1985
Modified Rankin Scale		I	n=100	++	N.A.	N.A.	VanSwieten 1988
		I	n=50	++	N.A.	N.A.	Wolfe 1992
		I	n=62	++	N.A.	N.A.	Wilson 2002
		I	n=438	N.A.	+++	N.A.	deHaan 1995
		I	n=59	N.A.	N.A.	++	Dromerick 2003
Beck Depression Inventory		var.	n=202	++			Aden 2002
		var.	N.A.	+++	+++	+	Salter 2005 a
Stroke Specific Quality of Life Scale	SSQOL	S	n=39	+++	++	N.A.	Williams 1999
Medical Outcome Study Short Form 36	SF-36	var.	n=2253	+++	++	+	Dorman 1999
		I	n=90	++	++	N.A.	Anderson 1998
Nottingham Health Profile	NHP	var.	n=53	++	N.A.	N.A.	Visser, 1995
			n=463	N.A.	++	+	Ebrahim 1993

Legende für die Administrationsform der Instrumente:

O = Beobachtung

I = Interview

S = Selbstausfüller-Fragebogen

var. = verschiedene Administrationsformen

Fehlende Bezeichnungen resultieren aus unvollständigen Angaben in den Quellenarbeiten.

4.3 Zusammenfassende Beurteilung von Skalen und Scores zur Schweregraddifferenzierung und Outcome Measurement bei Schlaganfall

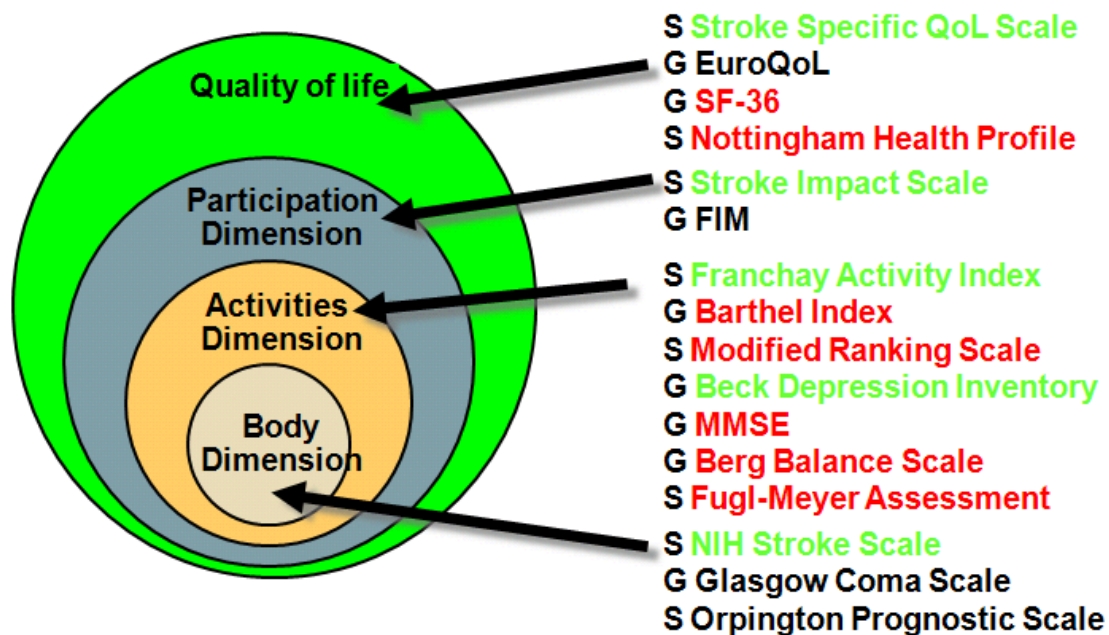


Abb. 4.3-1: Zusammenfassende Beurteilung der Skalen und Scores bei Schlaganfall

LEGENDE:

- G** generisches Instrument
- S** diagnosespezifisches Instrument
- rot** erhebliche Schwächen
- schwarz** geeignet
- Grün** empfohlen

Testgütekriterien zur Beurteilung herangezogen

Abb. 4.3-1 zeigt eine zusammengefasste Beurteilung von Skalen und Scores zur Schweregraddifferenzierung und Ergebnismessung bei Schlaganfall, der die im Abschnitt Methodik dargestellten Kriterien zugrunde liegen, innerhalb derer Reliabilität und Validität als Voraussetzungen attribuiert sind und ein besonderer Fokus auf die Änderungssensitivität gelegt wurde.

Die Abbildung zeigt, dass diagnosespezifische Instrumente die bessere Testgüte aufweisen als generische Instrumente. Diese Art der Messung bietet sich an, wenn tendenziell homogene Patient/innengruppen miteinander verglichen oder bewertet werden sollen. Auch generische Instrumente (das sind jene, die diagnoseübergreifend konzipiert und eingesetzt werden) sind für einen Einsatz in der Neurorehabilitation geeignet. Generische Instrumente werden i. d. R. dann eingesetzt, wenn Outcomes bei unterschiedlichen Diagnosegruppen (Erkrankungen) verglichen werden sollen.

Insofern ist die Entscheidung für ein Messinstrument nicht alleine von deren Testgüte (und der Machbarkeit und Akzeptanz der Messinstrumente) abhängig, sondern vom Zweck, den Messungen in Mess-Systemen dienen sollen. Erst wenn dieser Zweck eindeutig formuliert ist, ist es möglich, die modulare Zusammensetzung von generischen und spezifischen Instrumenten(teilen) zu entwerfen. Dies bedeutet, die Parameter festzulegen, die mit einem Kerninstrument erfasst werden können, welches nach Möglichkeit auch bei anderen Diagnosegruppen der Neurorehabilitation angewendet werden kann, und jene Parameter, die zusätzlich zum Kerninstrument diagnosespezifisch erfasst werden müssen, zu kombinieren.

**diagnosespezifische
Instrumente schneiden
besser bei homogenen
Patientengruppen ab,**

**generische Instrumente
bei unterschiedlichen
Diagnosegruppen**

5 Schädel-Hirn-Traumata (SHT)

5.1 Schweregrad differenzierung

Insgesamt wurde bis zum Jahr 2000 über mehr als 50 Instrumente publiziert, die Schädel-Hirn-Trauma-Patient/-innen anhand des Schweregrades der Erkrankung oder bezogen auf Behandlungsergebnisse klassifizieren.¹⁹⁶

in den letzten 10 Jahren
mehr als 50
Messinstrumente
entwickelt

5.1.1 Glasgow Coma Scale (GCS)

Die Glasgow Coma Scale/GCS ist das weitverbreitetste Instrument zur Schweregrad differenzierung bei Bewusstseinsstörungen. Die Messung basiert auf motorischen und verbalen Fähigkeiten und Augenaufschlag und misst Tiefe und Dauer komatöser Gesundheitszustände.¹⁷³ Sie wird nicht nur bei Schädel-Hirn-Traumatisierten, sondern auch bei einer Reihe anders verursachter Bewusstseinsstörungen eingesetzt. Die Skala ist einfach einsetzbar. Die Interrater Reliabilität ist teilweise schlecht: Neurochirurg/-innen variieren in ihren Einschätzungen bis zu 20 %, andere Chirurg/-innen und Pflegepersonal bis zu 28 %.¹⁷⁴ Die Fehler, die bei der Einschätzung gemacht werden können, sind so groß (teilweise mehr als 1 Punkt auf einer 5 Punkte Skala), dass Studien, die nur die Interrater Reliabilität evaluieren, ihren Einfluss nicht zur Gänze erfassen können. Am größten sind die Fehler bei mittleren Schweregraden. Gesundheitspersonal braucht also in der Regel Training für den Einsatz der GCS.¹⁷⁵

weitverbreitet,
Schweregrad-
differenzierung bei
Bewusstseinsstörungen

schlechte Interrater
Reliabilität

fehleranfällig
geringer prognostischer
Nutzen

Training notwendig

Die Fähigkeit der Glasgow Outcome Scale/GOS, Outcomes vorherzusehen, ist auch bei früher Anwendung gering.¹⁷⁶

5.1.2 Revised Trauma Score (RTS)

Der Revised Trauma Score (RTS) ist ein physiologisches Instrument, das aus der GCS, systolischem Blutdruck und der Atmungsfrequenz besteht. Zwei Items (capillary refill, respiratory extension) wurden hingegen exkludiert. Dadurch erreicht der RTS höhere Sensitivität, aber geringere Spezifität. Die Vorhersagekraft bezüglich Mortalität steigt.¹⁷⁷

GOS + Blutdruck +
Atmungsfrequenz = RTS

geringfügige
Verbesserungen

5.1.3 Trauma and Injury Severity Score (TRISS_{RTS})

Ein weiterentwickelter Subscore der RTS, der Trauma and Injury Severity Score (TRISS_{RTS}) hat sich als effektives Messinstrument für den anatomischen und physiologischen Schweregrad von Schädel-Hirn-Traumata erwiesen und ist international akzeptiert.¹⁷⁸ Der TRISS_{RTS} besteht aus der RTS, dem Injury Severity Score/ISS und dem Alter. Diese Komponenten korrelieren in hohem Maße und erlauben eine akzeptable Vorhersage der Mortalität und des Gesundheitszustandes bei Entlassung aus dem akutstationären Setting.¹⁷⁹

RTS+ISS+Alter=TRISS_{RTS}

hohe Testgüte

Die meisten Messsysteme der Schweregrad differenzierung wurden in den U.S.A. entwickelt und validiert. Zwischen U.S.A. und Österreich gibt es jedoch sowohl bei den Subtypen der Traumata als auch in der Organisation der Erstversorgung nach Unfällen und eventuell in den therapeutischen

Vergleichsstudie von
8 Instrumenten

Konzepten Unterschiede. Bevor diese Systeme in das Österreichische Setting übertragen werden, sind Validierungsstudien empfehlenswert. Eine solche, die prospektiv designte Kölner Validationsstudie untersuchte acht verschiedene Systeme der Schweregraddifferenzierung an einer Stichprobe von 612 Patient/innen:

Score	Sensitivity (%)	Specificity (%)	Accuracy (%)	Cut-off Point ^a
TRISS _{TS}	94.2	90.1	91.3	<0.88
TRISS _{RTS}	92.3	90.1	90.8	≤0.87
Prehospital Index	90.5	91.5	90.8	≥8
Trauma Score	88.9	90.5	90.8	≤11
Revised Trauma Score	87.3	90.1	89.2	≤5.4
Glasgow Coma Scale	87.8	89.1	88.7	≤6
Injury Severity Score	81.5	89.1	86.8	≥28
Polytraumaschlüssel	75.1	89.8	85.3	≥26

^a To allow for comparisons, the cut-off point was chosen in such a way that specificity was next to 90.0%.

Abb. 5.1-1: Testgüte verschiedener Instrumente der Schweregraddifferenzierung bei Schädel-Hirn-Traumata¹⁸⁰

**TRISS_{RTS} besteht Vergleich
am besten**

**Validität bestätigt, große
Akzeptanz der TRISS_{RTS}**

Diese Studie von Buillon¹⁸¹ hat gezeigt, dass alle getesteten Instrumente ihre Fähigkeit, Trauma-Patient/innen hinsichtlich ihres Schweregrades korrekt zu klassifizieren, unter Beweis stellen konnten. Die Autoren kommen zum Schluss, dass vier Gründe für den forcierten Einsatz des TRISS_{RTS} (für den Aufbau einer deutschlandweiten Datenbasis) sprechen: 1. TRISS_{RTS} sei international weithin akzeptiert, 2. im Vergleich zur gesamten TRISS-Score benötigt TRISS_{RTS} weniger Variablen zur Kalkulation, 3. die Validität wurde in Studien in den U.S.A., U.K., Australien und Deutschland bestätigt, der TRISS_{RTS} Score bezieht sinnvollerweise prognostische Variablen wie das Altern in die Kalkulation mit ein (Face Validität).

5.1.4 Schweregraddifferenzierung bei Kindern und Jugendlichen

**unterschiedliche
Unfallursachen**

**klassische Confounder:
Case mix**

**für Erwachsene
konzipierte
Messinstrumente zeigen
Schwächen &
Implementierungs-
schwierigkeiten basiert auf
validierten Instrumenten
für Erwachsene**

Die Ursachen von Unfällen im Kindes- und Jugendalter variiert nach dem Alter. Während Kleinkinder häufiger stürzen, werden Kinder im Vorschulalter eher von KFZ angefahren oder als Passagiere in Verkehrsunfälle involviert. Ältere Kinder stürzen primär mit dem Fahrrad oder werden als Fahrradfahrer in Verkehrsunfälle involviert. Obwohl verschiedene Ausprägungsformen der Unfallverletzungen resultieren, wird der Schweregrad der Verletzung des Gehirns als gemeinsamer Nenner herangezogen und prognostische Einschätzungen getroffen, die sich für gewöhnlich auf Mortalität und Morbidität (ICF-Klasse 1) beziehen. Jeder ergebnisorientierte Vergleich von Behandlungseinrichtungen wird vom unterschiedlichen Case-Mix verzerrt.

¹⁸²

Die Versuche, Messinstrumente die für Erwachsene konzipiert wurden auf Kinder und Jugendliche zu übertragen, scheiterten teilweise u.a. deshalb, weil sie ausschließlich auf Mortalität fokussierten und Morbidität sowie längerfristige Funktionsbeeinträchtigungen nicht berücksichtigten und zudem mathematische

Komplexität aufwiesen, die für das Behandlungspersonal schwierig handhabbar waren.

Pediatric Trauma Score (PTS)

Die für Kinder und Jugendliche (in der Evaluationsforschung, vorwiegend U.S.) am häufigsten eingesetzten Instrumente sind der Injury Severity Score (ISS), die Glasgow Coma Scale (GCS), der Paediatric Trauma Score (PTS) oder der Revised Trauma Score (RTS).

Eine spezielle Weiterentwicklung der Injury Severity Score (ISS), die New ISS (NISS) stellen Sullivan et al. vor.¹⁸³ Sowohl ISS als auch mit leichten Vorteilen die NISS eignen sich zur Schweregraddifferenzierung bei unter 20jährigen Patient/innen.

Tepas et al.¹⁸⁴ entwickelten und testeten den Pediatric Trauma Score (PTS) im Hinblick auf seine Testgüte zur Unterscheidung verschiedener Schweregrade von SHT bei Kindern an zwei Gruppen: Eine Gruppe von 110 Patient/innen wurde in einem Behandlungszentrum von einem einzelnen Mediziner bewertet, eine zweite Gruppe (matched) bestand aus 120 Patient/innen aus dem US National Pediatric Trauma Registry, deren Daten aus verschiedenen Behandlungszentren stammten. Die Studie dokumentiert die Konstruktvalidität der PTS im Vergleich mit dem ISS und ihre Reliabilität bei der Bestimmung verschiedener Schweregrade.

Weiterentwicklung: New Injury Severity Score

Pediatric Trauma Score

konstruktvalide

Pediatric Risk Indicator (PRI)

RTS und PTS wurden auch zur Triage eingesetzt. Der PTS wurde dabei als ohne Zusatznutzen im Vergleich zur RTS bringend sowie als praktisch zu komplex kritisiert, weshalb Tepas et al.¹⁸⁵ den Pediatric Risk Indicator (PRI) entwickelten. Dabei handelt es sich um eine relative einfach anwendbare mathematische Methode, die den Einsatz von ISS, GCS und PTS voraussetzt und als Verhältnis von $PRI = ISS / (GCS + PTS)$ kalkuliert wird. Die Formel berücksichtigt die Skalierungsrichtungen der einzusetzenden Instrumente und resultiert in $PTS < 1$ für jene Gruppe von Patient/innen, deren Schweregrad ernst zu nehmend sei. Die Autoren dokumentierten, dass dadurch eine signifikante Unterscheidung in Bezug auf Mortalität, Aufenthaltsdauer und FIM-Werten bei Entlassung gemacht werden kann.

Versuch, Komplexität zu reduzieren

5.1.5 Prognosepotential von Instrumenten der Schweregraddifferenzierung

Die Vorhersagekraft der Messinstrumente in Bezug auf die längerfristigen Ergebnisse der Behandlung gehört zu den bedeutsamsten Qualitätskriterien der Skalen. Die Prognose bestmöglich zu erfassen ist wichtig, um die (akutstationären) Rehabilitationsziele planen zu können, um die angebrachte postakute Rehabilitation zu planen, um das Behandlungspersonal in Bezug auf physiologische und psycho-soziale Behandlungsziele zu trainieren und nicht zuletzt um die Ressourcen bestmöglich zu allozieren.¹⁸⁶

für Therapieplanung wichtig

Mehrere Studien untersuchten, inwiefern die Summenwerte der Schweregraddifferenzierungsskalen mit der längerfristigen Funktionalität (outcomes) korrelieren. Persinger¹⁸⁷ untersuchte die Stärke des Zusammenhangs bei niedrigen und mittleren Schweregraden und bezog (wie die nachfolgenden genannten Studien auch) die Ergebnisse physiologischer Standard-Untersuchungen (beispielsweise Anomalien im EEG) in die Analysen mit ein. Die Ergebnisse werfen eine Reihe von Fragestellungen auf, so korrelier-

ten beispielsweise die Symptome des organischen Hirnsyndroms invers ($r = -0,35$) mit der Dauer des Komas, der posttraumatischen Amnesie und der Anzahl der EEG-Anomalien.

RTS und GCS inakzeptabel

Zafonte et al.¹⁸⁸ hatten bis dahin den RTS mit inakzeptabler Prädiktion des Gesundheitszustands bei Entlassung aus dem akutstationären Setting untersucht, in einer späteren Studie der Autoren erreicht auch die GCS nicht die erforderliche Testgüte. Zu denselben Ergebnissen kamen auch Hall et al.¹⁸⁹

Table 3. Disability Rating Scale Correlations

Variable	Correlation Coef.	p Value	N
GCS	-0.386	<0.0001	112
RTS	-0.394	<0.0001	109
TRISS	-0.458	<0.0001	109
ISS	0.34	0.0002	119
ALS Head and Neck	0.181	0.049	119
Age	0.171	0.0623	120

Table 3. Disability Rating Scale Correlations

From: Wagner: J Trauma, Volume 49(3).September 2000.411-419

Abb. 5.1-2: Testgütevergleich der Disability Rating Scale (Anm.: hier als Referenz für den „wahren prognostischen Wert“ mit anderen Skalen der Schweregrad differenzierung)

TRISS im Vergleich am genauesten, aber unzureichend

Modul aus diagnostischen Subskalen ineffektiv

Wagner et al. fanden heraus, dass die höchste Korrelation der Instrumente der Schweregrad differenzierung mit dem Outcome für die TRISS erreicht werden können, anzuzweifeln bleibt, ob ein Korrelationskoeffizient von $r = 0,46$ darauf schließen lässt, dass dies ausreichend genau möglich ist:

Beim Versuch, ein Messinstrument zu validieren, welches ausschließlich aus körperlichen Subskalen besteht (vgl. Functional Capacity Index), um die funktionale Entwicklung von Patient/innen in den Ergebnisdimensionen 2 und 3 der ICF zu prognostizieren, scheiterten Schluter et al.¹⁹⁰

5.2 Outcome Scales

5.2.1 Glasgow Outcome Scale (GOS)

weitverbreitet

Die Glasgow Outcome Scale/GOS gehört zu den in der Ergebnisforschung weitverbreiteten Instrumenten, die in der neurologischen Rehabilitation verwendet werden und besteht aus einem Item, anhand dessen Patient/innen in der Regel von Kliniker/innen bewertet werden. Erstmals wurde sie 1979, 2 Jahre nach der Glasgow Coma Scale, in der Literatur beschrieben.¹⁹¹ Bond & Janett differenzierten den Gesundheitszustand nach Schädel-Hirn-Traumata ursprünglich in 5 Kategorien: Tod; Apallisches Syndrom; Schwere Behinderung (bei Bewusstsein, aber behindert); Mäßige Behinderung (behindert, aber unabhängig); Gute Erholung. In einer erweiterten Version (GOS-Extended, GOS-E) werden die Behinderungsgrade jeweils zweigeteilt, sodass insgesamt 8 Kategorien entstehen.

Brooks et al.¹⁹² zeigten, dass die Inter-Rater-Reliabilität bei der Skala mit 3 Behinderungsgraden höher ist als bei der GOS-E, insgesamt aber nur bis zu 3 Monaten nach Diagnose zufriedenstellend. Ähnliche Werte erzielten spätere Weiterentwicklungen der GOS, wie etwa die Edinburgh Extended Glasgow Outcome Scale.¹⁹³

**Reliabilität
zufriedenstellend**

Auf weitere Klassifikationsprobleme bei Kindern macht eine deutsche Studie aufmerksam.¹⁹⁴ Den Ergebnissen zufolge leidet ein beträchtlicher Anteil (12-31 %) jener Kinder, deren Outcome auf der GOS 1 Jahr nach der Diagnose die höchste Kategorie erreicht, an mittelgradigen Hirndefiziten. In ähnlicher Weise verweisen die Ergebnisse einer weiteren Studie darauf, dass die GOS nicht mehr als 25 % der schwer beeinträchtigten Patient/innen identifizieren kann.¹⁹⁵ Satz et al.¹⁹⁶ untersuchten die Validität der Kategorien 3-5 GOS mit dem Ergebnis, dass 6 Monate nach der Diagnose die Kategorie 4 (mäßige Behinderung) nicht eindeutig zwischen schwerer Behinderung und guter Erholung unterscheiden konnte.

**Klassifikationsprobleme
bei Kindern**

**Ergebnisse einer
Validationsstudie**

Die GOS und GOS-E stehen auch deshalb unter Kritik, weil sie lediglich auf die körperliche Funktionalität und Behinderung abzielen und somit emotionale und soziale Selbstbewertung der Patient/innen ausschließt. Wilson et al.¹⁹⁷ haben diesbezüglich gezeigt, dass die Ergebnisse in allen Kategorien mit Depressionsskalen sowie dem SF-36 eindeutig und spezifisch, aber nur teilweise signifikant, korrelieren. Konkurrente Validität besteht ebenso im Vergleich zur Disability Rating Scale (DRS) und dem Barthel Index (BI).

**Eindimensional: nur
körperliche
Funktionalität**

aber valide

Anderson et al.¹⁹⁸ publizierten 1993 eine Studie über unterschiedliche Bewertungen durch Ärzt/innen, Psycholog/innen und die Angehörige. Psycholog/innen, die die Patient/innen neuropsychologisch getestet hatten, bewerteten die Patient/innen ähnlich Angehörigen, deren Bewertungen via postalischen Fragebogen eingeholt wurden ($r = 0.79$; $p = 0.001$). Die Bewertungen durch Allgemeinmediziner/innen, ebenfalls postalisch eingeholt, und Psycholog/innen differierten hingegen stärker ($r = 0.49$; $p = 0.001$). Die Ärzt/innen tendierten zur Überschätzung des Gesundheitszustands der Patient/innen. Die Autor/innen empfehlen, dass Patient/innen idealerweise von Personal, welches nicht in die akute Behandlung involviert war, bewertet werden sollten, alternativ von im Umgang mit der Skala trainiertem Behandlungspersonal. In vielen anderen Studien wird nicht transparent beschrieben, wo, von wem und wie die Interviews zur Einschätzung der GOS oder GOS-E Items beschrieben wurden.

**geringe Interrater
Reliabilität**

**Ärzt/innen tendieren zur
Überschätzung des
Gesundheitszustands**

5.2.2 Rivermead Post-Concussion Symptoms Questionnaire (RPQ)

Da bis dahin keine Instrumente zur Symptom-Einschätzung infolge Bewusstseinsstörungen entwickelt waren, veröffentlichten King et al. 1995 den Rivermead Post-Concussion Symptoms Questionnaire.¹⁹⁹ Seine Items sind mit „Ja“ oder „Nein“ zu beantworten, wodurch er eine hohe Interrater Reliabilität erzielt. Zehn Jahre später veröffentlichten Eyres et al. eine Validationsstudie des Instruments mit zweifelhaften Ergebnissen. Der RPQ erfüllt grundlegende psychometrische Validitätskriterien nur sehr mangelhaft, ist unidimensional und weist einen Floor-Effekt auf.²⁰⁰

**hohe Interrater
Reliabilität**

zweifelhafte Validität

5.2.3 Barthel Index (BI)

generisches Instrument

Der Barthel Index (BI) wurde ursprünglich entwickelt, um Fortschritte in der Selbstständigkeit und Mobilität während der Rehabilitation von Schlaganfall-Patient/innen zu messen und hat sich zu einem generisch eingesetzten Ergebnisinstrument der physiologischen Komponente der Funktionalität in der Rehabilitation entwickelt.²⁰¹ Es handelt sich um einen Index der Aktivitäten des täglichen Lebens.²⁰² Der BI ist in seiner Ausrichtung limitiert, da er psycho-soziale Funktionalität weitgehend ausspart. Reliabilität und Validität sind zwar für den Einsatz bei Schlaganfall-Patient/innen gegeben, in der Erforschung von Schädel-Hirn-Traumata hat sich der BI allerdings nicht bewährt. Zudem verhindert ein veritabler Ceiling-Effekt den Nachweis von Verbesserungen im Gesundheitszustand.²⁰³

Dimensionen limitiert:

körperliche
Funktionalität

für SHT nicht
empfehlenswert

5.2.4 Functional Independence Measure (FIM)

weitverbreitet

Das Functional Independence Measure/FIM wird seit Beginn der 1990er Jahre entwickelt und gehört insbesondere in den USA zu den weitverbreiteten Instrumenten zur Schweregrad differenzierung und Ergebnismessung in der Rehabilitationsforschung. Es besteht aus 18 Items und ist ordinalskaliert. Das FIM kann generisch für alle Diagnosegruppen in der Rehabilitation eingesetzt werden. Bei Schädel-Hirn-Traumata wird die Schwäche des Instruments, Ergebnisse der Interventionen nicht ausreichend diagnosespezifisch messen zu können, durch den Einsatz zusätzlicher Skalen wie beispielsweise dem Functional Assessment Measure/ FAM zu kompensieren versucht. Es dauert 20-30 min, um den Fragebogen gemeinsam mit Patient/innen, auf der Grundlage von Beobachtung oder im Rahmen eines telefonischen Interviews auszufüllen. Gemessen werden motorische und kognitive Funktionalität.

generisch

hohe Reliabilität, gute Validität

Dodds et al. wiesen an einer Studienpopulation von mehr als 11.000 Patient/innen exzellente Reliabilitätswerte und gute Validität des Instruments nach.²⁰⁴ Grey und Kennedy zeigten zeitgleich ausreichend hohe Korrelation zwischen der Bewertung durch Arzt/innen während der Rehabilitation und einer Selbstbewertung durch Patient/innen 6 Wochen nach Entlassung.²⁰⁵ Zur Frage, ob Bewertungen durch das Behandlungspersonal im Vergleich zu ausführlichem Befragen der Patient/innen durch das Personal zu unterschiedlichen Ergebnissen mit dem FIM führen, führten Karamehmetoglu et al.²⁰⁶ eine Studie durch, die bei erheblichen methodischen Schwächen – beide Bewertungsverfahren wurden konsekutiv von/m der/m gleichen Mediziner/in durchgeführt – keine gravierenden Unterschiede feststellen konnte.

verschieden administrierbar

Mortiffee et al. überprüften die Testgüte einer auf 3 Items reduzierten FIM-Skala mit Ergebnissen, die eine Item-Reduktion nicht befürworten.²⁰⁷

erheblicher Ceiling- Effekt

Hall et al.²⁰⁸ untersuchten das FIM sowie weitere in der Rehabilitation verwendete Schweregrads- und Ergebnisinstrumente bezüglich ihrer Ceiling-Effekte und fanden heraus, dass das FIM einen erheblichen Ceiling-Effekt (50 %) aufweist. Wenn es mit dem Functional Assessment Measure/FAM kombiniert eingesetzt und indexiert wird, ist der Ceiling-Effekt weiterhin sehr hoch, wenn auch auf 30 % reduziert. Die Disability Rating Scale/DRS zeigt hingegen weit weniger Ceiling-Effekt bei allen Mess-Zeitpunkten (bei Entlassung sowie ein und zwei Jahre danach). Die Community Integration Scale/CIS erwies sich in dieser Studie als unpräzise, vor allem was jene Be-

reiche sozialer Integration betrifft, die bei Schädel-Hirn-Traumata unter Mitleidenschaft gezogen werden. Die Studie folgert, dass es für Rehabilitation im ambulanten Setting exakterer Messinstrumente bedarf.

Zur Frage, ob Bewertungen durch das Behandlungspersonal im Vergleich zu ausführlichem Befragen der Patient/innen durch das Personal zu unterschiedlichen Ergebnissen mit dem FIM führen, führten Karamehmetoglou et al.²⁰⁹ ein Studie durch, die bei erheblichen methodischen Schwächen – beide Bewertungsverfahren wurden konsekutiv von/m der/m gleichen Mediziner/in durchgeführt – keine gravierenden Unterschiede feststellen konnte.

Patient/innen in verschiedenen Behandlungseinrichtungen können systematisch unterschiedlich eingeschätzt werden. Während es bei der Entlassung aus einem akut-stationären Behandlungssetting zur Überschätzung in der Kategorie „self care“ kommt, überschätzen in der Behandlungskette nachfolgende Behandlungseinrichtungen die Kategorie „soziale Kognition“. Insgesamt gibt es jedoch eine zufriedenstellende Übereinstimmung ($r=0,83$).²¹⁰

Lundgren-Nilsson et al.²¹¹ verglichen die Validität des FIM als generischem Instrument für drei verschiedene Diagnosegruppen (Schädel-Hirn-Trauma, Schlaganfall, Rückenmarksverletzungen) anhand des Rasch Modells und bemängelten, dass genauere Analysen erst nach einer Reihe von wiederholten Bewertungen vorgenommen werden konnten sowie dass das FIM vor allem bei der Messung von Behandlungsergebnissen von Rückenmarksverletzten Schwächen zeigte.

gute Reliabilität auch zwischen verschiedenen Behandlungskontexten

Rasch-Modell: hohe Dimensionalität

5.2.5 Functional Assessment Measure (FAM)

Das Functional Assessment Measure/FAM wird zusätzlich zum FIM eingesetzt, um jene Ergebniskategorien zu messen, die bei Schädel-Hirn-Traumata relevant sind und vom FIM nicht ausreichend abgebildet werden. Es besteht aus 12 Items in den Kategorien kognitive Funktionalität, Verhalten, Kommunikation und soziale Funktionalität. Die Skala mit insgesamt 30 Items wird als FIM+FAM bezeichnet und kann in etwa 30 min ausgefüllt werden.

Die internen Konsistenzen der Items, Subskalen und Skalen von FIM+FAM wurden an einer geeigneten Studienpopulation in der Rehabilitation nach SHT zufriedenstellend getestet.²¹²

Die Interrater Reliabilität ist beim Einsatz in der Rehabilitation nach Schädel-Hirn-Traumata adäquat bis ausgezeichnet, Schwächen in den psychosozialen Kategorien sind evident.²¹³ Auch McPherson et al.²¹⁴ berichten diese Ergebnisse in einer anderen Studie.

McPherson & Pentland²¹⁵ untersuchten die Testgütekriterien von FIM+FAM im Vergleich mit dem Barthel Index. Der Studie zufolge ist der Ceiling-Effekt des Barthel Index weitaus höher: Während keine/r der 54 Patient/innen laut Barthel-Index eine messbare Beeinträchtigung vorwies, erzielten weniger als 10 Patient/innen Bestwerte auf FIM+FAM. Da FIM+FAM eine größere Bandbreite an möglichen Beeinträchtigungen abbilden, erzielten sie auch eine höhere Konstruktvalidität und signifikantere Zusammenhänge der Ergebnisse.

**FIM+FAM zur Erhöhung der Spezifität
30 Items**

gute Reliabilität

Ceiling-Effekt des Barthel Index weitaus höher als FIM+FAM

**akzeptable
Vorhersagekraft der
Partizipation**

Gurka et al.²¹⁶ untersuchten in einer prospektiven Studie, inwiefern durch den Einsatz von FIM+FAM 6 und 24 Monate nach der Entlassung aus einem akutstationären Setting psychosoziale Ergebnisse wie die Wiederaufnahmefähigkeit einer Erwerbstätigkeit (gemessen anhand der „Return-to-work-Skala“) und soziale Integration (gemessen anhand des „Community Integration Questionnaire“) vorhergesagt werden können. Nach 6 Monaten zeigen sowohl FIM als auch FAM akzeptable Vorhersagequalitäten, nach 24 Monaten gelingt dies nur anhand der Kategorie Motor auf dem FAM. Die Autoren interpretieren das Zustandekommen der Ergebnisse durch Ceiling-Effekte und geringe „ökologische Validität“ der Instrumente.

5.2.6 Disability Rating Scale (DRS)

multidimensional

Die Disability Rating Scale/DRS wurde in einem akutstationären Rehabilitationssetting entwickelt, um Veränderungen der Funktionalität während der Rehabilitation zu messen. Das Instrument zeichnet sich durch seine Bandbreite an Outcomes aus, die gemessen werden kann (ICF-Klassen 1-3) und wird in der Literatur gemeinhin als einfach verwend- und administrierbar (2-15 min, Einschätzung durch Behandlungspersonal, Interview mit Angehörigen, Selbstausfüller) beschrieben. Es besteht aus 8 Items, die 4 Kategorien zugeordnet werden: Wachheit und Bewusstsein (1), kognitive Fähigkeiten zur Selbsthilfe (2), Pflegeabhängigkeit (3) und psychosoziale Integration (Arbeit, Haushalt, Schule) (4). Patient/innen können maximal 29 Punkte erzielen, was einem apallischen Syndrom in voller Ausprägung entspricht. Bei 0 Punkten bestehen keine Beeinträchtigungen. Die DRS darf nicht angewendet werden, wenn Patient/innen unter dem Einfluss von bewusstseinsverändernden Medikamenten stehen.

**ausgezeichnete
Interrater Reliabilität**

Die Interrater Reliabilität der DRS wurde von Rappaport et al. (1982) an einer Studienpopulation von 88 akutstationären Patient/innen mit $r=0.97$ errechnet. An einer kleineren aber vergleichbaren Population später von Gouvier²¹⁷ mit $r=0.98$ bestätigt, ebenso wie die Test-Retest-Reliabilität ($r=0.95$). Bei hohem Schweregrad der SHT gelangen Novack et al.²¹⁸ zum Schluss, dass sowohl bei Aufnahme als auch bei Entlassung aus dem Krankenhaus hohe Übereinstimmung von Familienangehörigen und Bewertungen durch Behandlungspersonal ($r=0.93-0.95$) erzielt werden kann.

konstruktvalide

Sowohl Rappaport et al.²¹⁹ als auch Hall et al.²²⁰ bestätigten die konkurrente Validität der DRS im Vergleich zu verschiedenen anderen Skalen. Gouvier verglich die DRS mit der GOS und fand Korrelationen von $r=0.5$ bei Aufnahme und $r=0.67$ bei Entlassung aus akutstationären Einrichtungen.

**hohe
Änderungssensitivität**

Hall et al.²²¹ verglichen die Nützlichkeit der DRS mit der GOS, um Veränderungen während des Krankenhausaufenthaltes zu messen, wobei sich die DRS als weit sensitiver (71 % vs. 33 %) hinsichtlich Veränderungen des Gesundheitszustands der Patient/innen erwies.

**Vorhersagekraft für
stationäre
Aufenthaltsdauer**

Ashley et al.²²² testeten die DRS im Hinblick auf ihre Eignung, Ergebnisse in der postakuten Rehabilitation nach SHT vorauszusagen. Das Instrument zeigt stabile Vorhersagewerte bis zu 14 Jahren nach der Entlassung, außer für die Wiedereingliederung in den Arbeitsmarkt.

Sowohl die Werte bei der akutstationären Aufnahme als auch bei Entlassung korrelierten signifikant mit der Wiederaufnahme der Erwerbsarbeit. Eine weitere Studie zeigt die Zuverlässigkeit der DRS-Aufnahmewerte zur Vorhersage der stationären Aufenthaltsdauer.²²³

Hinweise auf die Grenzen des Einsatzes der DRS zeigen Boake & High.²²⁴ Sie verglichen die Selbstbewertung der Patient/innen anhand der GOS und der DRS mit der Craig Handicap Assessment and Reporting Technique (CHART) und zeigten, dass die GOS und DRS-Skalen die sozial erwünschten Ergebnisse, vor allem Wiederaufnahme der Erwerbstätigkeit und soziale Integration nicht gut abbilden. In einer anderen Studie, die verschiedene Schweregrad- und Ergebnisskalen hinsichtlich ihrer Voraussagekraft für die Wiedereingliederung ins Berufsleben ein Jahr nach einem Unfall untersuchte, erwies sich die DRS neben anderen als zuverlässiges Instrument.²²⁵ Unter Berücksichtigung des Berufsstatus vor einem Unfall konnte die Wiederaufnahme der Erwerbstätigkeit in bis zu 66 % der Fälle richtig vorhergesagt werden.²²⁶

Vorhersagekraft der Partizipation

Zhang et al.²²⁷ verglichen zwei Instrumente, die die Partizipation („Handicap“) Schädel-Hirn-Traumatisierter erfassen – den „Community Integrations Questionnaire“ (CIQ) und die „Craig Handicap Assessment and Reporting Technique“ (CHART) mit der DRS. Die Autoren sehen den CIQ als exaktestes Instrument, um diese Dimension der ICF (WHO) zu erfassen. Im Vergleich zwischen DRS und CHART schneidet die DRS insofern besser ab, als dass die Items und Skalen, die Partizipation erfassen, stärker korrelieren als gemessen anhand der CHART.

DRS um Instrumente für Partizipation ergänzen: CIQ

Die Fülle an klinischen Studien, in der die DRS seit ihrer Einführung verwendet wurde, welche in diesem Report aber aufgrund der Eingrenzung der Fragestellung nicht einbezogen wurden, suggeriert, dass die DRS mittlerweile als „Gold Standard“ in der Evaluationsforschung der Rehabilitation bei Schädel-Hirn-Traumata gilt.

Gold Standard bei SHT

5.2.7 Functional Status Examination (FSE)

Ebenso wie die DRS wurde die Functional Status Examination/FSE entwickelt, um insbesondere Veränderungen der Funktionalität von Schädel-Hirn-traumatisierten Patient/innen verlässlich zu messen. Ebenso wie die DRS verfügt sie über die gesamte Bandbreite an Ergebnisdimensionen. Dikmen et al.²²⁸ testeten die FSE an 138 Patient/innen mit akzeptabler Test-Retest-Reliabilität ($r=0,8$) und ebenso hoher Korrelation zwischen einer Selbstaussfüllerversion und einem strukturierten Interview mit Angehörigen ($r=0,8$). Die Validität wurde aus weiter nicht nachvollziehbaren Gründen nicht anhand der Standards (FIM, DRS), sondern anhand der (zu diesem Zeitpunkt als vom Verfasser überholt eingeschätzten) GOS getestet. Hudak et al.²²⁹ verglichen die Skala mit der GOS-E und kamen zu ähnlichen Ergebnissen.

akzeptable Reliabilität

akzeptable Validität

5.2.8 Independent Living Scale (ILS)

Die Independent Living Scale (ILS) wurde über eine 15jährige Periode hinweg in einer postakuten Rehabilitationseinrichtung entwickelt und psychometrisch getestet.²³⁰ Operationalisiert wurden die Kategorien „Aktivitäten des täglichen Lebens“, „Verhalten“ und „Initiation“. Datenbasis waren 5250 Patient/innen zwischen 1986 und 1995. Das Instrument misst bei hohem methodischem Aufwand Unabhängigkeit im täglichen Leben mit sehr hoher Genauigkeit (Interrater Reliabilität: 0,85; Konvergente Validität: 0.82-0,87).

hohe Reliabilität und Validität, bei hohem Aufwand

stabile Vorhersagekraft

Ashley et al.²³¹ untersuchten dann auch die Stabilität der Ergebnisdimension „Independent Living“, die durch Rehabilitation bei jungen Erwachsenen eintritt, anhand der ADL. Mit Ausnahme der Dimension „berufliche Integration/ employability“, die in einem Zeitraum von 14 Jahren nach Entlassung rückläufig war. In Abhängigkeit von der Zeitspanne seit der Entlassung konnten Prognosewerte anhand der ILS entscheidend genauer eingeschätzt werden als aufgrund demographischer Daten.

5.2.9 Community Integration Questionnaire (CIQ)

misst soziale Integration

Der Community Integration Questionnaire (CIQ) misst soziale Integration in die Familie und Gesellschaft. In verschiedenen Zusammenhängen wird dieses Konzept der gesundheitsbezogenen Lebensqualität anhand von Informationen über (instrumentelle) Aktivitäten des täglichen Lebens, Hobbies und Produktivität im Haushalt, bei der Arbeit, in der Schule, im Zusammenleben in der Gemeinschaft, familiäre und berufliche Integration in Netzwerke sowie in soziale Rollenfunktionalität operationalisiert.

nur durchschnittliche Reliabilität,

Interrater Reliabilität gut untersucht

Dijkers²³² untersuchte die Testgütekriterien des CIQ in klinischen Studien. Die Interrater Reliabilität ist mit Korrelationskoeffizienten zwischen 0,4 und 0,6 durchschnittlich gut, später korrigierten Tepper et al.²³³ die Testgüte des Instruments hinsichtlich seiner Reliabilität jedoch nach unten, wenn Angehörige anstelle der Patient/innen selbst interviewt werden. Jene Dimension der Skala, die Integration zu Hause messen, funktioniert nicht zufriedenstellend. Sander et al. finden in einer anderen Studie, dass einzelne Kategorien wie „Schule“ zu $r=0.94$ korrelieren.²³⁴

konstruktvalide

Die interne Konsistenz des Instruments ist in den Studien von Willer & Ottenbacher²³⁵ und Corrigan²³⁶ höher als 0,8. Sander et al. testeten verschiedene Formen der Validität des Community Integration Questionnaire an 316 Patient/innen mit vielen häufig als „Gold Standards“ bezeichneten Instrumenten und bestätigten dessen Konstruktvalidität.²³⁷

Ceiling-Effekt-Gefahr

Dennoch sind mit dem Instrument erhebliche Schwierigkeiten verbunden, so sind beispielsweise viele der Items in der Allgemeinbevölkerung nicht normalverteilt. Formale Validitätsstudien wurden bisher kaum durchgeführt, in der Literatur wird über Ceiling-Effekte einzelner Dimensionen spekuliert.²³⁸

änderungssensitiv

gute Unterscheidung zwischen Patient/innen mit/ohne ambulante Rehabilitation

Nichtsdestotrotz belegen weitere Studien das Potential des Instruments zur sensitiven Erfassung möglicher Veränderungen der sozialen Integration. In der Studie von Willer & Ottenbacher²³⁹ gelingt es, anhand des CIQ trennscharf zwischen Patient/innen, die zuhause weitgehend unabhängig leben, Patient/innen mit begrenzter Hilfe aus der Nachbarschaft und Patient/innen in medizinischen Einrichtungen zu unterscheiden. Willer et al. berichten über die Änderungssensitivität der CIQ, dass anhand der Summenwerte die erwarteten Fortschritte bei Patient/innen mit und ohne ambulante Rehabilitation unterschieden werden konnten.²⁴⁰

Kuipers et al. verglichen den Einsatz von CIQ und Sydney Psychosocial Reintegration Scale/SPRS an 96 Patient/innen und 121 Vergleichspersonen aus der Allgemeinbevölkerung, die sich hinsichtlich Match-Kriterien nicht signifikant unterschieden, mit den Ergebnissen, dass die SPRS leichte psychometrische Vorteile mit sich bringt, was auch auf unterschiedliche Rücklaufzeiten bei fehlenden Werten zurückzuführen sein könnte, die sich als Resultat einer Online-Befragung ergaben.²⁴¹

Die Integration in soziale Netzwerke und das gesellschaftliche Leben als Ergebnisse zu messen, stellt nach wie vor ein teilweise offenes Problem dar. In den U.S.A. wurden dazu ein Konsensus-Prozess geführt, den Sherwin et al. darstellen.²⁴²

offene Probleme

Johnston et al.²⁴³ entgegneten dem Problem, in dem sie die Struktur der Follow-Up-Werte von 231 Patient/innen in der "US National TBI Model Systems Database" mit der Rasch-Analysemethode analysierten. Unter allen in diesem System gemessenen globalen Outcomes inklusive Funktionalität, Aktivität und Partizipation (ICF-Kategorien 1-3) führte die Dimensionalität der Instrumente bei über 90 % zu sehr guten Ergebnissen. Bei etwa 10 % war die Funktionalität anhand der eingesetzten Instrumente (FIM, NFI) nicht eindeutig messbar, was daran lag, dass verschiedene Instrumente modular eingesetzt wurden, deren Dimensionen nicht zueinander passten. Insbesondere neuropsychologische Item-Batterien, aber auch auf die gesellschaftliche Partizipation abstellende Instrumente funktionierten unzureichend. Die Autoren schlagen ein eindimensionales Mess-System vor, das sie als „general community functioning“ bezeichnen.

Erfahrung in Datapools

Bedeutung der ICF-Multidimensionalität

5.3 Patient/innen-Zufriedenheit

Wasser et al.²⁴⁴ entwickelten und testeten einen Fragebogen zur Patient/innen-Zufriedenheit, der Informationen der Angehörigen bezüglich der Behandlung auf Intensivstationen in 20 Items und 5 Kategorien erhebt und aus den üblichen Kriterien der Patient/innen-Zufriedenheit besteht. Reliabilität und Validität einer anderen Skala, der „Family Satisfaction Scale“ (FSS) wurde von Underhill et al.²⁴⁵ spezifisch mit den Angehörigen von SHT betroffenen Patient/innen getestet, in dem sie die FSS bei ausgezeichneten Reliabilitätswerten 12 und 60 Monate nach Entlassung einsetzten.

validiertes Instrument für das akutstationäre Intensivsetting

validierte Instrumente für SHT-Patient/innen

5.4 Exkurs: Rückenmarksverletzungen

Lam et al.²⁴⁶ verfassten 2008 ein systematisches Review, um jene Ergebnisinstrumente bezüglich ihrer Testgüte gegenüberzustellen, die in der Rehabilitation von Rückenmarksverletzungen angewendet werden können. Insgesamt wurden 7 Instrumente identifiziert und in Time-Walking-Tests und kategorische Instrumente differenziert. Die Time-Walking-Tests verfügen über exzellente Reliabilität, Validität und Änderungssensitivität. Die psychometrische Beschaffenheit kategorischer Instrumente unterschied sich weit stärker und nur jene Instrumente, die spezifisch für Rückenmarksverletzungen entwickelt wurden, verfügten über überdurchschnittliche Testgüte. Das FIM war demnach bei Rückenmarksverletzten nicht verlässlich und nicht valide, während diagnosespezifische Instrumente wie das Spinal Cord Independence Measure/SCIM sich hier zum „Gold Standard“ entwickelten.

Review zu Instrumenten bei Rückenmarksverletzungen

FIM nicht verlässlich

Spinal Cord Independence Measure/SCIM als "gold standard"

Tab. 5.4-1: Übersicht der Skalen und Scores bei Schädel-Hirn-Traumata

Instrument	Bez.	ICFclass	Admin	pop	Reliabilität	Validität	Änderungs- sensitivität	Quelle
Glasgow Outcome Scale	GOS	1,2	var.	n=58	++	N.A.	N.A.	Anderson 1993
				n=59	N.A.	+	N.A.	Koelfen 1997
				I n=113	N.A.	++	N.A.	Wilson 2000
Glasgow O S extended	GOS-E	1,2	var.	n=57	+	+	N.A.	Brooks 1986
Edinburgh extended GOS	EEGOS	1,2	var.		++	N.A.	N.A.	Hellawell 1997
Rivermead Post Concussion Symptoms Questionnaire		1	var.	n=46	++	N.A.	N.A:	King 1995
			var.	n=369	+	+	N.A.	Eyres 2005
Functional Independence Measure	FIM	1,2		n=11102	+++	++	+	Dodds 1993
				n=14799	N.A.	N.A.	++	Karamehmetoglu 1997
				n=230	N.A.	N.A.	++	Long 1997
				n=57	++	N.A.	N.A.	Segal 1993
FIM limited set		1,2		n=83	-	-	-	Mortifee 1996
Functional Assessment Measure	FIM+FAM	1,2	rated	n=54	++	N.A.	N.A.	Donaghy, 1998
			rated	n= 88	N.A.	++	+	Gurka, 1999
			rated	n=965	++	+++	N.A.	Hawley 1999
			rated	n=54	N.A.	++	++	McPherson 1997
Disability Rating Scale	DRS	1,2,3	var.	n=40	+++	+++	N.A.	Gouver 1987
				n=88	+++	++	+	Rappaport 1982
				n=332	N.A.	N.A.	++	Ashley 1997

Instrument	Bez.	ICFclass	Admin	pop	Reliabilität	Validität	Änderungs- sensitivität	Quelle
				n=132	N.A.	N.A.	++	Cifu 1997
				n=184	N.A.	N.A.	++	Eliason 1984
				n=70	N.A.	N.A.	+++	Hall 1985
Functional Independence Level	FIL	1,2	S	n=338	++	N.A.	N.A.	Schatz 2002
Functional Status Examination	FSE	1,2	I	n=138	++	+	N.A.	Dikmen 2001
Independent Living Scale		2,3	I		+++	+++	+++	Ashley 2001
Community Integration Quest. Revisited		2,3	var.	n=315	N.A.	N.A.	+	Ashley 1997
			I	n=322	++	++	N.A.	Kuijpers 2004
			I	n= 312	++	+++	N.A.	Sander 1999
Brain Injury Community Rehabilitation Outcome	BICRO	2,3	S,I		+	+	N.A.	Powell 1998
Functional Capacity Index ¹	FCI	1	TA	n=617	N.A.	N.A.	- (adverse)	Schluter 2005
Pediatric Risk Index*	PRI	1,2	S	n=5345	++.	++	N.A.	Tepas 1997
Critical Care Family Satisfaction Survey		N.A.	S	n=237	++	++	N.A.	Wasser 2001
Family Satisfaction Scale	FSS	N.A.	S		+++	+	N.A.	Underhill 2004
Readyness to Change Questionnaire	RTC	N.A.	I		N.A.	++	N.A.	Bombardier 2000

5.5 Zusammenfassende Beurteilung von Skalen und Scores zur Schweregraddifferenzierung und Outcome measurement bei Schädel-Hirn-Traumata

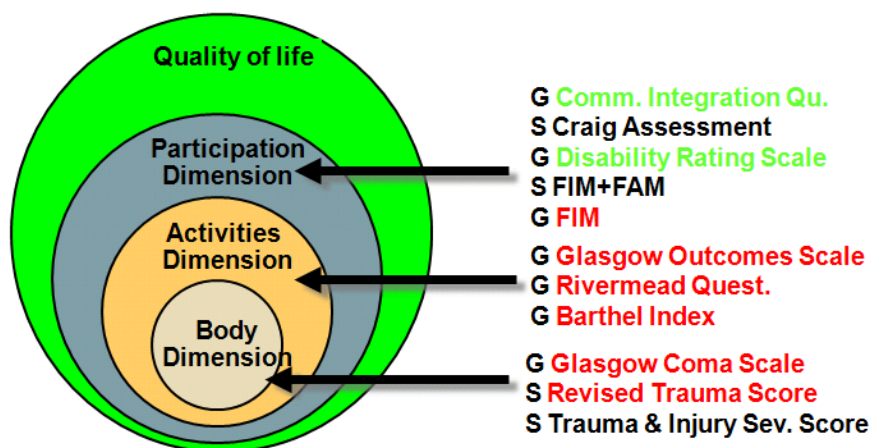


Abb. 5.5-1: Zusammenfassende Beurteilung der Skalen und Scores bei Schädel-Hirn-Trauma

LEGENDE:

G	generisches Instrument
S	diagnosespezifisches Instrument
rot	erhebliche Schwächen
schwarz	geeignet
Grün	empfohlen

Testgütekriterien zur Beurteilung herangezogen

diagnosespezifische Instrumente nicht notwendigerweise die bessere Testgüte generische Instrumente u.U. besser

Abb. 5.5-1 zeigt eine zusammengefasste Beurteilung von Skalen und Scores zur Schweregraddifferenzierung und Outcome Measurement bei Schädel-Hirn-Trauma, der die im Abschnitt Methodik dargestellten Kriterien zugrunde liegen, innerhalb derer Reliabilität und Validität als Voraussetzungen attribuiert sind und ein besonderer Fokus auf die Änderungssensitivität gelegt wurde.

Die Abbildung zeigt, dass diagnosespezifische Instrumente im Vergleich nicht notwendigerweise die bessere Testgüte aufweisen. Diese Art der Messung bietet sich an, wenn tendenziell homogene Patient/innengruppen miteinander verglichen werden sollen. Spezifische Instrumente sind bei Schädel-Hirn-Trauma unterentwickelt. Die vorhandenen generischen Instrumente (das sind jene, die diagnoseübergreifend konzipiert und eingesetzt werden) sind für einen Einsatz in der Neurorehabilitation unter Umständen

besser geeignet. Generische Instrumente werden i. d. R. dann eingesetzt, wenn Outcomes bei unterschiedlichen Diagnosegruppen (Erkrankungen) verglichen werden sollen. Insofern ist die Entscheidung für ein Messinstrument nicht alleine von deren Testgüte (und der Machbarkeit der Messungen) abhängig, sondern vom Zweck, den Messungen in Mess-Systemen dienen sollen. Erst wenn dieser Zweck eindeutig formuliert ist, ist es möglich, die modulare Zusammensetzung von generischen und spezifischen Instrumenten(teilen) zu designen. Dies bedeutet, die Parameter festzulegen, die mit einem Kerninstrument erfasst werden können, welches nach Möglichkeit auch bei anderen Diagnosegruppen der Neurorehabilitation angewendet werden kann, und jene Parameter, die zusätzlich zum Kerninstrument diagnosespezifisch erfasst werden müssen, zu kombinieren.

6 Ausblick

Skalen und Scores sind verschiedenen Spannungsfeldern ausgesetzt, die sich lt. Masur²⁴⁷ auch als Pole beschreiben lassen, bei denen sich eine Annäherung an den einen Pol zwangsläufig einer Entfernung vom anderen Pol entspricht. Diese Spannungsfelder können nach unterschiedlichen Begriffspaaren differenziert werden.

- ✿ Kollektivistisch versus individuell: Ein Test, der ein Kollektiv gut charakterisiert, kann individuelle Eigenheiten nur schlecht darstellen.
- ✿ Generalisierbar versus sensitiv: Die Validität eines Instruments, welches die Outcomes mehrerer Diagnosen oder von Diagnosegruppen der Neurorehabilitation abbildet, leidet in Bezug auf die Sensitivität der Messung von Behandlungsergebnissen in den Untergruppen.
- ✿ Objektiv versus relevant: Eine Skala mit hoher Objektivität, z.B. eine Skala zur Erfassung einer Beeinträchtigung besitzt in einem größeren Bezugsrahmen (etwa in Bezug auf die Veränderung des Lebensstils) u. U. wenig Relevanz.
- ✿ Analytisch vs. holistisch: Wenn man nur Teile betrachtet, erzielt man in Bezug auf einen Teilaspekt eine hohe Objektivität und eine hohe Sensitivität. In Bezug auf das Ganze kann man jedoch völlig fehlgeleitet sein.

Bei Überlegungen zur Auswahl von Messsystemen zur Unterstützung der neurologischen Rehabilitation in Österreich muss

- ✿ Zunächst die Zielsetzung der Messungen klar sein, sowie
- ✿ ob und in welcher Form Messungen zur Qualitätssicherung und Ergebnismessung in der Neurorehabilitation eingesetzt werden sollen,
- ✿ ob und in welcher Form Messungen mit der Refundierung und den entsprechenden ökonomischen Anreizsystemen kombiniert werden sollen,

Darüberhinaus gilt es in einem nächsten Schritt zu eruieren,

- ✿ wie sich die jeweiligen Vorgaben auf die Auswahl der Messinstrumente auswirken und
- ✿ wie die Qualität der Messungen unter den jeweiligen Vorgaben sichergestellt werden kann.

**Messinstrumente
befinden sich im
Spannungsfeld zwischen**

Polen:

**kollektivistisch vs.
individuell,**

**generalisierbar vs.
sensitiv,**

etc.

Quellen

-
- ¹ Bak P., Lohsträter A., Müller W.D., Bocker B., Smolenski U.C.: Zertifizierungs- und Akkreditierungssysteme als Instrument des Qualitätsmanagements in der Rehabilitation (Teil 2) – Charakterisierung der meist verbreiteten Systeme. *Phys Rehab Med Kuror* 2004; 14 (6): 283-290.
- ² Bak P., Schmidt J., Lohsträter A., Müller W. D. : Beurteilung der funktionalen Gesundheit in der Rehabilitation von Patienten mit Verletzungen. *Phys Med Rehab Kuror* 2005; 15: 83-89.
- ³ Farin E., Glattacker M., Follert P., Kuhl C., Jackel W. H.: Effekte und Prädiktoren des Outcomes in der Rehabilitation muskuloskeletaler Erkrankungen. *Phys Rehab Med Kuror* 2004; 14 (3): 123-133.
- ⁴ Gerdes N., Best S, Jackel W.H.: Akutbehandlung – Frührehabilitation – Anschlussrehabilitation – Vorschlag zu einem 3-Phasen-Modell bei Implantation einer Endoprothese an Hüfte oder Knie. *Phys Rehab Med Kuror* 2004; 14 (4): 179-186.
- ⁵ Farin E., Engel E.M., Dimou A., Jackel W.H.: Die Erwartungen von Rehabilitations- und Vorsorgeeinrichtungen an externe Qualitätssicherungsmaßnahmen. *Phy Rehab Med Kuror* 2004; 14 (2): 73-81.
- ⁶ Linn P. L., Slinde J. A.: Determination oft he significance of change between pre- and posttesting periods. *Reviews of Educational Research* 1977; 47: 121-150. Zit. n. : Steiner D. L. & Norman G. R. (1995): *Health measurement scales. A practical guide to their development and use. Second Edition.* Oxford University Press. 163-180.
- ⁷ Guyatt G., Walter S., Norman G.: Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; 40: 171–178.
- ⁸ Deyo R. A., Centor R. M.: Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986; 39: 897–906.
- ⁹ Kazis L. E., Anderson J. J., Meenan R. F.: Effects sizes for interpreting changes in health status. *Med Care* 1989; 27: 178–189.
- ¹⁰ Steiner D. L. & Norman G. R. (1995): *Health measurement scales. A practical guide to their development and use. Second Edition.* Oxford University Press. 65.
- ¹¹ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹² Frytak J.: Measurement. *Journal of Rehabilitation Outcomes Measurement* 2000; 4: 15-31. Zit. n. Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹³ Dijkers M., Kropp G. C., Esper R. M., Yavuzer G. et al.: Reporting on reliability and validity of outcome measures in medical rehabilitation research. *Disability and Rehabilitation* 2002; 24: 819-827.

-
- ¹⁴ Jutai J., Teasell R.: The necessity and limitations of evidence based practice in stroke rehabilitation. *Topics in Stroke Rehabilitation* 2003; 10: 71-78.
Zit. n. Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹⁵ Weltgesundheitsorganisation (Hrsg.): Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit. Deutsches Institut für Medizinische Dokumentation und Information - DIMDI. 2005. WHO, Genf
- ¹⁶ World Health Organization (2002): Towards a Common Language for Functioning, Disability and Health: ICF. Geneva, World Health Organization;
- ¹⁷ Brenner M. H., Curbow B., Legro M. W.: The proximal-distal continuum of multiple health outcome measures: the case of cataract surgery. *Medical Care* 1995; 33: 236-244.
- ¹⁸ Steiner D. L. & Norman G. R. (1995): Health measurement scales. A practical guide to their development and use. Second Edition. Oxford University Press. 508.
- ¹⁹ Fitzpatrick R., Davey C., Buxton M. J., Jones D. R.: Evaluation of patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2: 1-74.
- ²⁰ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ²¹ Josephson S. A., Hills N. K., Johnston S. C.: NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovascular Diseases* 2006; 22(5-6):389-395.
- ²² Masur H. (2000): Skalen und Scores in der Neurologie. Stuttgart: Thieme Verlag. 212.
- ²³ Meyer B. C., Hemmen T. M., Jackson C. M., Lyden P. D.: Modified National Institutes of Health Stroke Scale for use in stroke clinical trials: prospective reliability and validity. *Stroke* 2002; 33(5):1261-1266.
- ²⁴ de Caneda M. A., Fernandes J. G., de Almeida A. G., Mugnol F. E.: Reliability of neurological assessment scales in patients with stroke. [Portuguese] [English Abstract]. *Arquivos de Neuro-Psiquiatria* 2006; 64(3A):690-697.
- ²⁵ Berger K., Weltermann B., Kolominsky-Rabas P., Meves S., et al.: The reliability of stroke scales. The german version of NIHSS, ESS and Rankin scales. [German] *Fortschritte der Neurologie-Psychiatrie* 1999; 67(2):81-93.
- ²⁶ Kasner S. E., Cucchiara B. L., McGarvey M. L., Luciano J. M. et al.: Modified National Institutes of Health Stroke Scale can be estimated from medical records. *Stroke* 2003; 34(2): 568-570.
- ²⁷ Kasner S. E., Chalela J. A., Luciano J. M., Cucchiara B. L. et al.: Reliability and validity of estimating the NIH stroke scale score from medical records. *Stroke* 1999; 30(8):1534-1537.

-
- ²⁸ Williams L. S., Yilmaz E. Y., Lopez-Yunez A. M.: Retrospective assessment of initial stroke severity with the NIH Stroke Scale. *Stroke* 2000; 31(4):858-862.
- ²⁹ Muir K. W., Weir C. J., Murray G. D., Povey C. et al.: Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke* 1996; 27(10):1817-1820.
- ³⁰ Cheung C. M., Tsoi T. H., Hon S. F., Au-Yeung M. et al.: Using the National Institutes of Health Stroke Scale (NIHSS) to predict the mortality and outcome of patients with intracerebral haemorrhage. *Hong Kong Medical Journal* 2008; 14(5):367-370.
- ³¹ Rieck M., Moreland J.: The Orpington Prognostic Scale for patients with stroke: reliability and pilot predictive data for discharge destination and therapeutic services. *Disability & Rehabilitation* 2006; 27(23): 1425-1433.
- ³² Studenski S. A., Wallace D., Duncan P. W., Rymer M. et al.: Predicting stroke recovery: three- and six-month rates of patient-centered functional outcomes based on the orpington prognostic scale. *Journal of the American Geriatrics Society* 2001; 49(3):308-312.
- ³³ Shoemaker M. J., Mullins-MacRitchie M., Bennett J., Vryhof K. et al.: Predicting response to rehabilitation in elderly patients with stroke using the Orpington Prognostic Scale and selected clinical variables. *Journal of Geriatric Physical Therapy* 2006; 29(2):69-73.
- ³⁴ Celik C., Aksel J., Karaoglan B.: Comparison of the Orpington Prognostic Scale (OPS) and the National Institutes of Health Stroke Scale (NIHSS) for the prediction of the functional status of patients with stroke. *Disability & Rehabilitation* 2006; 28(10):609-612.
- ³⁵ Wright C. J., Swinton L. C., Green T. L., Hill M. D.: Predicting final disposition after stroke using the Orpington Prognostic Score. *Canadian Journal of Neurological Sciences* 2004; 31(4):494-498.
- ³⁶ Muir K. W., Weir C. J., Murray G. D., Povey C. et al.: Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke* 1996; 27(10):1817-1820.
- ³⁷ Kalra L., Crome P.: The role of prognostic scores in targeting stroke rehabilitation in elderly patients. *Journal of the American Geriatrics Society* 1993; 41(4):396-400.
- ³⁸ Teasdale G., Jennett B.: Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; 2(7872): 81-84.
- ³⁹ Teasdale G., Knill-Jones R., van der Sande J.: Observer variability in assessing impaired consciousness and coma. *J Neurol Neurosurg Psychiatry* 1978; 41(7): 603-610.
- ⁴⁰ Rowley G., Fielding K.: Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet* 1991; 337: 535-538.
- ⁴¹ Weir C. J., Bradford A. P., Lees K. R.: The prognostic value of the components of the Glasgow Coma Scale following acute stroke. *Qjm* 2003; 96(1):67-74.

-
- ⁴² Weingarten S., Bolus R., Riedinger M. S., Maldonado L. et al.: The principle of parsimony: Glasgow Coma Scale score predicts mortality as well as the APACHE II score for stroke patients. *Stroke* 1990; 21(9):1280-1282.
- ⁴³ Counsell C., Dennis M., McDowall M.: Predicting functional outcome in acute stroke: comparison of a simple six variable model with other predictive systems and informal clinical prediction. *Journal of Neurology, Neurosurgery & Psychiatry* 2004; 75(3):401-405.
- ⁴⁴ Mahoney F., Barthel D.: Functional evaluation: the Barthel Index. *Md State Med J.* 1965; 14:61-65.
- ⁴⁵ Banks, J. L., Marotta, C. A.: Outcomes Validity and Reliability of the Modified Rankin Scale: Implications for Stroke Clinical Trials: A Literature Review and Synthesis. *Stroke* 2006; 38(3): 1091-1096.
- ⁴⁶ Van Swieten J. C., Koudstaal P. J., Visser M. C. , Schouten H. et al.: Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1988; 19: 604-607.
- ⁴⁷ Wolfe C. D., Taub N. A., Woodrow E. J., Burney P. G.: Assessment of scales of disability and handicap for stroke patients. *Stroke* 1991; 22: 1242-1244.
- ⁴⁸ Wilson J. T., Hareendran A., Grant M., Baird T. et al.: Improving the assessment of outcomes in stroke: Use of structured interview to assign grades on the Modified Rankin Scale. *Stroke* 2002; 33: 2243-2246.
- ⁴⁹ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Activity. *Disability and Rehabilitation* 2005; 27(6): 315-340.
- ⁵⁰ Wilson J. T., Hareendran A., Grant M., Baird T. et al.: Improving the assessment of outcomes in stroke: Use of structured interview to assign grades on the Modified Rankin Scale. *Stroke* 2002; 33: 2243-2246.
- ⁵¹ de Haan R. J., Limburg M., Bossuyt P., van der Meulen J et al.: The clinical meaning of Rankin 'handicap' grades after stroke. *Stroke* 1995; 26: 2027-2030.
- ⁵² Dromerick A. W., Edwards D. F., Diringer M. N.: Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials. *Journal of Rehabilitation Research and Development* 2003; 40: 1-8.
- ⁵³ Banks, J. L., Marotta, C. A.: Outcomes Validity and Reliability of the Modified Rankin Scale: Implications for Stroke Clinical Trials: A Literature Review and Synthesis. *Stroke* 2006; 38(3): 1091-1096.
- ⁵⁴ Banks J. L., Marotta C. A.: Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. *Stroke* 2007; 38(3): 1091-1096.
- ⁵⁵ Horner R. D., Swanson J. W., Bosworth H. B., Matchar D.B.: Effects of race and poverty on the process and outcome of inpatient rehabilitation services among stroke patients. *Stroke* 2003; 34: 1027-1031.
- ⁵⁶ Masiero S., Avesani R., Armani M., Verena P. et al.: Predictive factors for ambulation in stroke patients in the rehabilitation setting: a multivariate analysis. *Clinical Neurology & Neurosurgery* 2007; 109(9):763-769.

- ⁵⁷ Huybrechts K. F., Caro J. J.: The Barthel Index and modified Rankin Scale as prognostic tools for long-term outcomes after stroke: a qualitative review of the literature. *Current Medical Research & Opinion* 2007; 23(7): 1627-1636.
- ⁵⁸ Roberts L., Counsell C.: Assessment of clinical outcomes in acute stroke trials. *Stroke* 1998; 29: 986-991.
- ⁵⁹ Beck A. T., Ware C. H., Mendelson M., Mock J et al.: An inventory for measuring depression. *Archives of General Psychiatry* 1961; 4:561-571.
- ⁶⁰ Aben I., Verhey F., Lousberg R., Lodder J. et al.: Validity of the Beck Depression Inventory, Hospital Anxiety and Depression Scale, SCL-90 and Hamilton Depression Rating Scale as screening instruments for depression in stroke patients. *Psychosomatics* 2002; 43:386-393.
- ⁶¹ Aben I., Verhey F., Lousberg R., Lodder J. et al.: Validity of the Beck Depression Inventory, Hospital Anxiety and Depression Scale, SCL-90, and Hamilton Depression Rating Scale as Screening Instruments for Depression in Stroke Patients. *Psychosomatics* 2002; 43: 5.
- ⁶² Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Body Functions. *Disability and Rehabilitation* 2005; 27(4): 191-207.
- ⁶³ Duncan P. W., Jorgensen H. S., Wade D. T.: Outcome measures in acute stroke trials. A systematic review and some recommendations to improve practice. *Stroke* 2000; 31: 1429-1438.
- ⁶⁴ Duncan P. W., Wallace D., Lai S. M., Johnson D. et al.: The Stroke Impact Scale version 2.0. Evaluation of reliability, validity and sensitivity to change. *Stroke* 1999; 30: 2131-2140.
- ⁶⁵ Duncan P. W., Bode R. K., Lai S. M., Perera S.: Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation* 2003; 84: 950-963.
- ⁶⁶ Duncan P. W., Jorgensen H. S., Wade D. T.: Outcome measures in acute stroke trials. A systematic review and some recommendations to improve practice. *Stroke* 2000; 31: 1429-1438.
- ⁶⁷ Finch E., Brooks D., Stratford P. W., Mayo N. E.: Physical Rehabilitations Outcome Measures. A Guide to Enhanced Clinical Decision-Making. Toronto, Ontario: Canadian Physiotherapy Association; 2002
- ⁶⁸ Duncan P. W., Bode R. K., Min Lai S., Perera S.: Glycine Antagonist in Neuroprotection Americans Investigators. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Archives of Physical Medicine & Rehabilitation* 2003; 84(7): 950-963.
- ⁶⁹ Duncan P. W., Lai S. M., Bode R. K., Perera S.: Stroke Impact Scale-16: A brief assessment of physical function. *Neurology* 2003; 60(2): 291-296.
- ⁷⁰ Duncan P. W., Wallace D., Lai S. M., Johnson D. et al.: The Stroke Impact Scale version 2.0. Evaluation of reliability, validity and sensitivity to change. *Stroke* 1999; 30: 2131-2140.
- ⁷¹ Edwards B., O'Connell B.: Internal consistency and validity of the Stroke Impact Scale 2.0 (SIS 2.0) and SIS-16 in an Australian sample. *Quality of Life Research* 2003; 12: 1135.

-
- ⁷² Edwards B., O'Connell B.: Internal consistency and validity of the Stroke Impact Scale 2.0 (SIS 2.0) and SIS-16 in an Australian sample. *Quality of Life Research* 2003; 12(8): 1127-1135.
- ⁷³ Duncan P. W., Lai S. M., Tyler D., Perera S. et al.: Evaluation of proxy responses to the Stroke Impact Scale. *Stroke* 2002; 33(11): 2593-2599.
- ⁷⁴ Duncan P. W., Reker D. M., Horner R. D., Samsa G. P. et al.: Performance of a mail-administered version of a stroke-specific outcome measure, the Stroke Impact Scale. *Clinical Rehabilitation* 2002; 16(5): 493-505.
- ⁷⁵ Kwon S., Duncan P., Studenski S., Perera S. et al.: Measuring stroke impact with SIS: construct validity of SIS telephone administration. *Quality of Life Research* 2006; 15(3): 367-376.
- ⁷⁶ Petersen C., Morfeld M., Bullinger M.: Testing and validation of the German version of the Stroke Impact Scale. [German] [English Abstract.] *Fortschritte der Neurologie-Psychiatrie* 2001; 69(6): 284-290.
- ⁷⁷ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Activity. *Disability and Rehabilitation* 2005; 27(6): 315-340.
- ⁷⁸ McDowell I., Newell C.: Measuring Health. A Guide to Rating Scales and Questionnaires. New York: Oxford University Press, 1996.
- ⁷⁹ Collin C., Wade D. T., Davies S., Horne V.: The Barthel ADL Index: a reliability study. *International Disabilities Studies* 1988; 10: 61-63.
- ⁸⁰ D'Olhaberriague L., Litvan I., Mitsias P., Mansbach H.: A reappraisal of reliability and validity studies in stroke. *Stroke* 1996; 27: 2331-2336.
- ⁸¹ Hsueh I. P., Lee M. M., Hsieh C. L.: Psychometric characteristics of the Barthel activities of daily living index in stroke patients. *Journal of the Formosa Medical Association* 2001; 100: 526-532.
- ⁸² Yeo D., Faleiro R., Lincoln N. B.: Barthel ADL Index: a comparison of administration methods. *Clinical Rehabilitation* 1995; 9: 34-39.
- ⁸³ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Activity. *Disability and Rehabilitation* 2005; 27(6): 315-340.
- ⁸⁴ Green J., Forster A., Young J. A.: Test-retest reliability study of the Barthel Index, the Rivermead Mobility Index, the Nottingham Extended Activities of Daily Living Scale and the Frenchay Activities Index in stroke patients. *Disability & Rehabilitation* 2001; 23(15): 670-676.
- ⁸⁵ Cohen M. I., Marino R. J.: The tools of disability outcomes research functional status measures. *Archives of Physical Medicine and Rehabilitation* 2000; 81: S21-29.
- ⁸⁶ Fortinsky R. H., Granger C. V., Seltzer G. B.: The use of functional assessment in understanding home care needs. *Medical Care* 1981; 19: 489-497.
- ⁸⁷ Shah S., Cooper B., Maas F.: The Barthel Index and ADL evaluation in stroke rehabilitation in Australia, Japan, the US and the USA. *The Australian Occupational Therapy Journal* 1992; 39: 5-13.

- ⁸⁸ Duncan P. W., Samsa G., Weinberger M., Goldstein L. et al.: Health status of individuals with mild stroke. *Stroke* 1997; 28: 740-745.
- ⁸⁹ Wood-Dauphinee S., Williams J. I., Shapiro S. H.: Examining outcome measures in a clinical study of stroke. *Stroke* 1990; 21: 731-739.
- ⁹⁰ Wood-Dauphinee S., Berg K. O., Bravo G., Williams J. L.: The Balance Scale: responsiveness to clinically meaningful changes. *Canadian Journal of Rehabilitation* 1997; 10: 35-50.
- ⁹¹ Stone S. P., Ali B., Auberleek I., Thompsell A. et al.: The Barthel Index in clinical practice: use on a rehabilitation ward for elderly people. *Journal of the Royal College of Physicians of London* 1994; 28: 419-423.
- ⁹² Hocking C., Williams M., Broad J., Baskett J.: Sensitivity of Shah, Vanclay and Cooper's Modified Barthel Index. *Clinical Rehabilitation* 1999; 13:141-147.
- ⁹³ Dromerick A. W., Edwards D. F., Diringer M. N.: Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials. *Journal of Rehabilitation Research and Development* 2003; 40: 1-8.
- ⁹⁴ Hsueh I. P., Lee M. M., Hsieh C. L.: Psychometric characteristics of the Barthel activities of daily living index in stroke patients. *Journal of the Formosa Medical Association* 2001; 100: 526-532.
- ⁹⁵ Wallace D., Duncan P. W., Lai S. M.: Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *Journal of Clinical Epidemiology* 2002; 55: 922-928.
- ⁹⁶ Hobart J. C., Thompson A. J.: The five item Barthel Index. *Journal of Neurology, Neurosurgery and Psychiatry* 2001; 71: 225-230.
- ⁹⁷ Hsueh I. P., Lin J. H., Jeng J. S., Hsieh C. L.: Comparison of the psychometric characteristics of the Functional Independence Measure, 5-item Barthel index and 10-item Barthel Index in patients with stroke. *Journal of Neurology, Neurosurgery and Psychiatry* 2002; 73: 188-190.
- ⁹⁸ Duncan P. W., Samsa G., Weinberger M., Goldstein L. et al.: Health status of individuals with mild stroke. *Stroke* 1997; 28: 740-745.
- ⁹⁹ van der Putten J. M. F., Hobart J. C., Freeman J. A., Thompson A. J.: Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery and Psychiatry* 1999; 66: 480-484.
- ¹⁰⁰ Collin C., Wade D. T., Davies S., Horne V.: The Barthel ADL Index: a reliability study. *International Disabilities Studies* 1988; 10: 61-63.
- ¹⁰¹ McDowell I., Newell C. :Measuring Health. A Guide to Rating Scales and Questionnaires. New York: Oxford University Press, 1996. zit. n. Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Activity. Disability and Rehabilitation 2005; 27(6): 315-340.
- ¹⁰² Linacre J. M., Heinemann A. W., Wright B. D., Granger C. V. et al.: The structure and stability of the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation* 1994; 75: 127-132

-
- ¹⁰³ Dodds TA, Martin DP, Stolov WC, Deyo RA.: A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil* 1993; 74(5): 531-536.
- ¹⁰⁴ Oczkowski W, Barreca S. The Functional Independence Measure: Its use to identify rehabilitation needs in stroke survivors. *Archives of Physical Medicine and Rehabilitation* 1993; 74: 1291-1294.
- ¹⁰⁵ Granger C. V., Cotter A. C., Hamilton B. B., Fiedler R. C.: Functional Assessment Scales: a study of persons after stroke. *Archives of Physical Medicine and Rehabilitation* 1993; 74: 133-138.
- ¹⁰⁶ Brock J. A., Goldie P. A., Greenwood K. M.: Evaluating the effectiveness of stroke rehabilitation: choosing a discriminative measure. *Archives of Physical Medicine and Rehabilitation* 2002; 83: 92-99.
- ¹⁰⁷ Dodds TA, Martin DP, Stolov WC, Deyo RA.: A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil* 1993; 74(5): 531-536.
- ¹⁰⁸ Ring H., Feder M., Schwartz J., Samuels G.: Functional measures of first-stroke rehabilitation inpatients: Usefulness of the Functional Independence Measure total score with a clinical rationale. *Archives of Physical Medicine and Rehabilitation* 1997; 78: 630-635.
- ¹⁰⁹ Dodds TA, Martin DP, Stolov WC, Deyo RA.: A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil* 1993; 74(5): 531-536.
- ¹¹⁰ Brock J. A., Goldie P. A., Greenwood K. M.: Evaluating the effectiveness of stroke rehabilitation: choosing a discriminative measure. *Archives of Physical Medicine and Rehabilitation* 2002; 83: 92-99.
- ¹¹¹ van der Putten J. M. F., Hobart J. C., Freeman J. A., Thompson A. J.: Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery and Psychiatry* 1999; 66: 480-484.
- ¹¹² Wallace D., Duncan P. W., Lai S. M.M Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *Journal of Clinical Epidemiology* 2002; 55: 922-928.
- ¹¹³ Dromerick A. W., Edwards D. F., Diring M. N.: Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials. *Journal of Rehabilitation Research and Development* 2003; 40: 1-8.
- ¹¹⁴ van der Putten J. M. F., Hobart J. C., Freeman J. A., Thompson A. J.: Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery and Psychiatry* 1999; 66: 480-484.
- ¹¹⁵ Schepers V. P., Ketelaar M., Visser-Meily J. M., Dekker J. et al.: Responsiveness of functional health status measures frequently used in stroke research. *Disability & Rehabilitation* 2006; 28(17): 1035-1040.
- ¹¹⁶ Holbrook M., Skilbeck C. E.: An activities index for use with stroke patients. *Age and Aging* 1983; 12: 166-170.

- ¹¹⁷ Wade D. T., Leigh-Smith J., Langton-Hewer R.: Social activities after stroke: measurement and natural history using the Frenchay Activities Index. *International Rehabilitation Medicine* 1985; 1: 176-181.
- ¹¹⁸ Wyller T. B., Sveen U., Bautz-Holter E.: The Frenchay Activities Index in stroke patients: agreement between scores by patients and by relatives. *Disability and Rehabilitation* 1996; 18: 454-459.
- ¹¹⁹ Piercy M., Carter J., Mant J., Wade D. T.: Inter-rater reliability of the Frenchay Activities Index in patients with stroke and their carers. *Clinical Rehabilitation* 2000; 14: 433-440.
- ¹²⁰ Schuling J., de Haan R. J., Limburg M., Groenier K. H.: The Frenchay Activities Index; assessment of functional status in stroke patients. *Stroke* 1993; 24: 1173-1177.
- ¹²¹ Poole J. L., Whitney S. L.: Assessments of motor function post stroke: A review. *Physical and Occupational Therapy in Geriatrics* 2001; 19: 1-22.
- ¹²² Dick J. P. R., Builoff R. J., Stewart A., Blackstock J. et al.: Mini-mental State Examination in neurological patients. *Journal of Neurology, Neurosurgery and Psychiatry* 1984; 47: 499.
- ¹²³ Tombaugh T. N., McIntyre N. J.: The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatric Society* 1992; 40: 922-935.
- ¹²⁴ Teng E. L., Chui H. C.: The modified Mini-Mental State (3MS) Examination. *Journal of Clinical Psychiatry* 1987; 48: 314-318.
- ¹²⁵ Grace J., Nadler J. D., White D. A., Guilmette T. J. et al.: Folstein vs. Modified Mini-Mental State Examination in geriatric stroke. *Archives of Neurology* 1995; 52: 477-484.
- ¹²⁶ Ozdemir F., Birtane M., Tabatabaei R., Ekuklu G. et al.: Cognitive evaluation and functional outcome after stroke. *American Journal of Physical Medicine and Rehabilitation* 2001; 80: 410-415.
- ¹²⁷ Lorentz W. J., Scanlan J. M., Borson S.: Brief screening tests for dementia. *Canadian Journal of Psychiatry* 2002; 47: 723-733.
- ¹²⁸ Suhr J. A., Grace J.: Brief cognitive screening of right hemisphere stroke: relation to functional outcome. *Archives of Physical Medicine and Rehabilitation* 1999; 80: 773-776.
- ¹²⁹ Gutachten von Dr. Peter Wipfler zur Erstfassung der vorliegenden Arbeit, S. 2
- ¹³⁰ Berg K. O., Wood-Dauphinee S., Williams J. L., Maki B.: Measuring balance in the elderly: preliminary development of an instrument. *Physiotherapy Canada* 1989; 41: 304-311.
- ¹³¹ Berg K. O., Wood-Dauphinee S., Williams J. L., Maki B.: Measuring balance in the elderly: validation of an instrument. *Canadian Journal of Public Health* 1992; 83: 7-11.
- ¹³² Blum L., Korner-Bitensky N.: Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. *Physical Therapy* 2008; 88(5):559-566.

-
- ¹³³ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Activity. *Disability and Rehabilitation* 2005; 27(6): 315-340.
- ¹³⁴ Wood-Dauphinee S., Berg K. O., Bravo G., Williams J. L.: The Balance Scale: responsiveness to clinically meaningful changes. *Canadian Journal of Rehabilitation* 1997; 10: 35-50.
- ¹³⁵ Mao H. F., Hsueh I. P., Tang P. F., Sheu C. F. et al.: Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke* 2002; 33: 1022-1027.
- ¹³⁶ Duncan P. W., Goldstein L. B., Horner R. D., Landsman P. B.: Similar motor recovery of upper and lower extremities after stroke. *Stroke* 1994; 25: 1181-1188.
- ¹³⁷ Gladstone D. J., Danells C. J., Black S. E.: The Fugl-Meyer Assessment of Motor Recovery after Stroke: A critical review of its measurement properties. *Neurorehabilitation and Neural Repair* 2002; 16: 232-240.
- ¹³⁸ Ware J. E., Sherbourne C. D.: The MOS 36-item short-form health survey (SF36) I: Conceptual framework and item selection. *Medical Care* 1992; 30: 473-483.
- ¹³⁹ McDowell I., Newell C.: Measuring Health. A Guide to Rating Scales and Questionnaires. New York: Oxford University Press; 1996.
- ¹⁴⁰ Andresen E. M., Meyers A. R.: Health-related quality of life outcomes measures. *Archives of Physical Medicine and Rehabilitation* 2000; 81: 30-45.
- ¹⁴¹ Segal M. E., Schall R. R.: Determining functional/health status and its relation to disability in stroke survivors. *Stroke* 1994; 25: 2391-2397.
- ¹⁴² Dorman P. J., Slaterry J., Farrell B., Dennis M., Sandercock P.: Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. *Stroke* 1998; 29: 63-68.
- ¹⁴³ Anderson C., Laubscher S., Burns R.: Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke* 1996; 27: 1812-1816.
- ¹⁴⁴ Brazier J. E., Walters S. J., Nicholl J. P., Kohler B.: Using the SF-36 and the Euroqol on an elderly population. *Quality of Life Research* 1996; 5: 195-204.
- ¹⁴⁵ Hayes V., Morris J., Wolfe C., Morgan M.: The SF-36 Health Survey Questionnaire: Is it suitable for use with older adults? *Age and Aging* 1995; 24: 120-125.
- ¹⁴⁶ Brazier J. E., Harper R., Jones N. M., O'Cathain A. et al.: Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British Medical Journal* 1992; 305: 160-164.
- ¹⁴⁷ O'Mahoney P. G., Rodgers H., Thomson R. G., Dobson R. et al.: Is the SF-36 suitable for assessing health status of older stroke patients? *Age and Aging* 1998; 27: 19 – 23;
- ¹⁴⁸ Hunt S. M., McEwen J., McKenna S.P.: Measuring health status: a new tool for clinicians and epidemiologists. *The Journal of the Royal College of General Practitioners* 1985; 35: 185-188.

-
- ¹⁴⁹ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹⁵⁰ Anderson C., Laubscher S., Burns R.: Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke* 1996; 27: 1812-1816.
- ¹⁵¹ Bohannon RW, Maljanian R, Landes M.: Test-retest reliability of short form (SF)-12 component scores of patients with stroke. *International Journal of Rehabilitation Research* 2004; 27(2): 149-150.
- ¹⁵² King J. T., Horowitz M. B., Kassam A. B., Yonas H. et al.: The short form-12 and the measurement of health status in patients with cerebral aneurysms: performance, validity, and reliability. *Journal of Neurosurgery* 2005; 102(3): 489-494.
- ¹⁵³ Muller-Nordhorn J., Nolte CH., Rossnagel K., Jungehulsing G. J. et al.: The use of the 12-item short-form health status instrument in a longitudinal study of patients with stroke and transient ischaemic attack. *Neuroepidemiology* 2005; 24(4): 196-202.
- ¹⁵⁴ Williams L. S., Weinberger M., Harris M. D., Clark D. O. et al.: Development of a stroke-specific quality of life scale. *Stroke* 1999; 30: 1362-1369.
- ¹⁵⁵ Williams L. S., Weinberger M., Harris M. D., Clark D. O. et al.: Development of a stroke-specific quality of life scale. *Stroke* 1999; 30:1362 – 1369.
- ¹⁵⁶ Czechowsky D., Hill M. D.: Neurological outcome and quality of life after stroke due to vertebral artery dissection. *Cerebrovascular Disease* 2002; 13: 192-197.
- ¹⁵⁷ The EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199-208.
- ¹⁵⁸ (Salter 2005 c, S. 512)
- ¹⁵⁹ Essink-Bot M. L., Drabbe P. F., Bonsel G. J., Aaronson N. K.: An empirical comparison of four generic health status measures: The Nottingham Health Profile, the Medical Outcomes Study 36-item Short Form Health Survey, the COOP/WONCA Charts, and the EuroQol instrument. *Medical Care* 1997; 35: 522-537.
- ¹⁶⁰ Brazier J. E., Walters S. J., Nicholl J. P., Kohler B.: Using the SF-36 and the Euroqol on an elderly population. *Quality of Life Research* 1996; 5:195-204.
- ¹⁶¹ Coast J., Peters T. J., Richards S. H., Gunnell D. J.: Use of the EuroQol among elderly acute care patients. *Quality of Life Research* 1998; 7: 1-10.
- ¹⁶² Dorman P. J., Waddell F., Slaterry J., Dennis M. et al.: Is the EuroQol a valid measure of health-related quality of life after stroke? *Stroke* 1997; 28(10): 1876-1882.
- ¹⁶³ Hunt S. M., McEwen J., McKenna S. P.: Measuring health status: a new tool for clinicians and epidemiologists. *The Journal of the Royal College of General Practitioners* 1985; 35: 185-188.

-
- ¹⁶⁴ Kohlmann T., Bullinger M., Kirchberger-Blumstein I.: German version of the Nottingham Health Profile (NHP): translation and psychometric validation. [German] *Sozial- und Präventivmedizin* 1997; 42(3): 175-185.
- ¹⁶⁵ Hunt S. M., McKenna S. P., McEwen J.: *The Nottingham Health Profile: User's Manual*. Manchester, England; 1989.
- ¹⁶⁶ Hunt S. M., McEwen J., McKenna S. P.: Measuring health status: a new tool for clinicians and epidemiologists. *The Journal of the Royal College of General Practitioners* 1985; 35: 185-188.
- ¹⁶⁷ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹⁶⁸ Visser M. C., Koudstaal P. J., Erdman R. A., Deckers J. W. et al.: Measuring quality of life in patients with myocardial infarction or stroke: a feasibility study of four questionnaires in The Netherlands. *Journal of Epidemiology and Community Health* 1995; 49: 513-517.
- ¹⁶⁹ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹⁷⁰ Ebrahim S., Barer D., Nouri F.: Use of the Nottingham Health Profile with patients after a stroke. *Journal of Epidemiology and Community Health* 1986; 40: 166-169.
- ¹⁷¹ Salter K., Jutai J. W., Teasell R., Foley N. C. et al.: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability and Rehabilitation* 2005; 27(9): 507-528.
- ¹⁷² Ahmed S., Mayo N. E., Corbiere M., Wood-Dauphinee S. et al.: Change in quality of life of people with stroke over time: true change or response shift? *Quality of Life Research* 2005; 14(3): 611-627.
- ¹⁷³ Teasdale G., Jennett B.: Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; 2(7872): 81-84.
- ¹⁷⁴ Teasdale G., Knill-Jones R., van der Sande J.: Observer variability in assessing impaired consciousness and coma. *J Neurol Neurosurg Psychiatry* 1978; 41(7): 603-610.
- ¹⁷⁵ Rowley G., Fielding K.: Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet* 1991; 337: 535-538.
- ¹⁷⁶ Waxman K., Sundine M. J., Young R. F.: Is early prediction of outcome in severe head injury possible? *Arch Surg* 1991; 126: 1237-1242.
- ¹⁷⁷ Champion H. R., Sacco W. J., Copes W. S., Gann D. S. et al.: A revision of the Trauma Score. *Journal of Trauma-Injury Infection & Critical Care*. 1989; 29(5): 623-629.
- ¹⁷⁸ Wagner A. K., Sasser H. C., Hammond F., Wierciszewski, D. et al.: Intentional Traumatic Brain Injury: Epidemiology, Risk Factors, and Associations with Injury Severity and Mortality. *J Trauma Inj Inf Critic Care* 2000; 49(3): 404-410.

¹⁷⁹ Bouillon B., Lefering R., Vorweg M., Tiling T. et al. : Trauma Score Systems: Cologne Validation Study. *J Trauma Inj Inf Critic Care* 1997; 42(4): 652-658.

¹⁸⁰ Bouillon B., Lefering R., Vorweg M., Tiling T. et al. : Trauma Score Systems: Cologne Validation Study. *J Trauma Inj Inf Critic Care* 1997; 42(4): 652-658.

¹⁸¹ Bouillon B., Lefering R., Vorweg M., Tiling, T.: Trauma Score Systems: Cologne Validation Study. *J of Trauma, Injury, Infection & Critical Care* 1997; 42(4): 652-658.

¹⁸² Tepas J. J., Veldenz H. C., Discala C., Pieper P.: Pediatric Risk Indicator: An Objective Measurement of Childhood Injury Severity. *J Trauma Inj Inf Critic Care* 1997; 43(2): 258-262.

¹⁸³ Sullivan Th., Haider A., Stephen M., DiRusso M. et al.: Prediction of Mortality in Pediatric Trauma Patients: New Injury Severity Score Outperforms Injury Severity Score in the Severely Injured. *J Trauma* 2003; 55: 1083-1088.

¹⁸⁴ Tepas J. J., Mollitt D. L., Talbert J. L., Bryant M.: The pediatric trauma score as a predictor of injury severity in the injured child. *Journal of Pediatric Surgery* 1987; 22(1): 14-8.

¹⁸⁵ Tepas J. J., Veldenz H. C., Discala C., Pieper P.: Pediatric Risk Indicator: An Objective Measurement of Childhood Injury Severity. *J Trauma Inj Inf Critic Care* 1997; 43(2): 258-262.

¹⁸⁶ Hanks R. A., Millis S. R., Ricker J. H., Giacino J. T. et al.: The Predictive Validity of a Brief Inpatient Neuropsychologic Battery for Persons With Traumatic Brain Injury. *Arch Phys Med Rehabil* 2008; 89: 950-957.

¹⁸⁷ Persinger M. A.: Clinical neurological indicators are only moderately correlated with quantitative neuropsychological test scores in patients who display mild-moderate brain impairment following closed-head injuries. *Perceptual & Motor Skills* 1995; 81(2): 1283-1292.

¹⁸⁸ Zafonte R. D., Hammond F. M., Mann N. R., Wood D. L. et al.: Revised trauma score: an additive predictor of disability following traumatic brain injury? *Am J Phys Med Rehabil* 1996; 75: 456-461.

¹⁸⁹ Hall K. M., Hamilton B. B., Gordon W. A., Zasler N.: Characteristics and comparisons of functional assessment indices: disability rating scale, functional independence measure, and functional assessment measure. *J Head Trauma Rehabil* 1993; 8: 60-74.

¹⁹⁰ Schluter Ph., Neale R., Scott D., Luchter St. et al.: Validating the Functional Capacity Index: A Comparison of predicted versus observed Total Body Scores. *J Trauma* 2005; 58: 252-263.

¹⁹¹ Bond M. R., Jennett W. B. et al.: The nature of physical, mental and social deficits contributing to the categories of good recovery, moderate and severe

disability in the Glasgow Global Outcome Scale. *Acta Neurochir Suppl* 1979; 28(1): 126-127.

¹⁹² Brooks D. N., Hosie J., Bond M. R., Jennett B., Aughton M.: Cognitive sequelae of severe head injury in relation to the Glasgow Outcome Scale. *Journal of Neurology, Neurosurgery, and Psychiatry* 1986; 49: 549-553.

¹⁹³ Hellawell D. J., Signorini D. F.: The Edinburgh Extended Glasgow Outcome Scale (EEGOS): rationale and pilot studies. *Int J Rehabil Res.* 1997; 20(4): 345-354.

¹⁹⁴ Koelfen W., Freund M., Dinter D., Schmidt B. et al.: Long-term follow up of children with head injuries-classified as "good recovery" using the Glasgow Outcome Scale: neurological, neuropsychological and magnetic resonance imaging results. *Eur J Pediatr.* 1997; 156(3): 230-235.

¹⁹⁵ León-Carrión J., Alarcón J. C., Revuelta M., Murillo-Cabezas F. et al.: Executive functioning as outcome in patients after traumatic brain injury. *Int J Neurosci* 1998; 94(1-2): 75-83.

¹⁹⁶ Satz P., Zaucha K., Forney D. L., McCleary C. et al.: Neuropsychological, psychosocial and vocational correlates of the Glasgow Outcome Scale at 6 months post-injury: a study of moderate to severe traumatic brain injury patients. *Brain Inj* 1998; 12(7): 555-567.

¹⁹⁷ Wilson J. T., Pettigrew L. E., Teasdale G. M.: Emotional and cognitive consequences of head injury in relation to the Glasgow outcome scale. *J Neurol Neurosurg Psychiatry* 2001; 70(2): 267-268.

¹⁹⁸ Anderson S. I., Housley A. M., Jones P. A., Slattery J. et al.: Glasgow Outcome Scale: an inter-rater reliability study. *Brain Inj.* 1993; 7(4): 309-17.

¹⁹⁹ King N. S., Crawford F., Wenden F. J., Moss N. E., Wade D. T.: The Rivermead Post Concussion Symptoms Questionnaire. A measure of symptoms commonly experienced after head injury and its reliability. *J Neurology* 1995; 242(9): 587-592.

²⁰⁰ Eyres S., Carey A., Gilworth G., Neumann V. et al.: Construct Validity and Reliability of the Rivermead Post Concussion Symptoms Questionnaire. *Clinical Rehabilitation* 2005; 19(8): 878-887.

²⁰¹ Wade D. T. (1996): Measurement in neurological rehabilitation. New York: Oxford University Press.

²⁰² Collin C., Wade D. T., Davis S. et al.: The Barthel ADL Index: a reliability study. *International Disability Studies* 1988; 10: 61-63.

²⁰³ Heyink J.: The Barthel Index. In: Hutchinson A., Bentzen N. and König-Zahn C. (ed.) (1997): Cross Cultural Health Outcome Assessment: a User's Guide. Ruinen: Barrier. 99-103.

²⁰⁴ Dodds T. A., Martin D. P., Stolov W. C., Deyo R. A.: A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil* 1993; 74(5):531-536.

²⁰⁵ Grey N., Kennedy P.: The Functional Independence Measure: a comparative study of clinician and self ratings. *Paraplegia* 1993; 31(7): 457-461.

²⁰⁶ Karamehmetoğlu S. S., Karacan I., Elbaşı N., Demirel G. et al.: The functional independence measure in spinal cord injured patients: comparison of questioning with observational rating. *Spinal Cord* 1997; 35(1): 22-25.

-
- ²⁰⁷ Mortifee P. R., Busser J. R., Anton H. A.: The performance of a limited set of items from the Functional Independence Measure for use in acute trauma care and rehabilitation. *Arch Phys Med Rehabil* 1996; 77: 436-439.
- ²⁰⁸ Hall K. M., Mann N., High W., Wright J. et al.: Functional measures after traumatic brain injury: ceiling effects of FIM, FIM+FAM, DRS and CIQ. *J Head Trauma Rehabil* 1996; 11(5): 27-39.
- ²⁰⁹ Karamehmetoğlu S. S., Karacan I., Elbaşı N., Demirel G. et al.: The functional independence measure in spinal cord injured patients: comparison of questioning with observational rating. *Spinal Cord* 1997; 35(1): 22-25.
- ²¹⁰ Segal M. E., Ditunno J. F., Staas W. E.: Interinstitutional agreement of individual functional independence measure (FIM) items measured at two sites on one sample of SCI patients. *Paraplegia* 1993; 31(10):622-631.
- ²¹¹ Lundgren-Nilsson A., Tennant A., Grimby G., Sunnerhagen K. S.: Cross-diagnostic validity in a generic instrument: an example from the Functional Independence Measure in Scandinavia. *Health and Quality of Life Outcomes* 2006; 4:1-8.
- ²¹² Hawley C. A., Taylor R., Hellawell D. J., Pentland B.: Use of the functional assessment measure (FIM+FAM) in head injury rehabilitation: a Psychometric analysis. *J Neurol Neurosurg Psychiatry* 1999; 67: 749-754.
- ²¹³ Donaghy S., Wass P. J.: Interrater reliability of the Functional Assessment Measure in a brain injury rehabilitation program. *Arch Phys Med Rehabil* 1998; 79(10): 1231-1236.
- ²¹⁴ McPherson K. M., Pentland B., Cudmore S. F., Prescott R. J.: An Interrater Reliability Study of the Functional Assessment Measure (FIM+FAM). *Disability & Rehabilitation* 1996; 18(7): 341-347.
- ²¹⁵ McPherson K. M., Pentland B.: Disability in patients following traumatic brain injury--which measure? *Int J Rehabil Res.* 1997; 20(1): 1-10.)
- ²¹⁶ Gurka J. A., Felmingham K. L., Baguley I. J., Schotte D. E. et al.: Utility of the functional assessment measure after discharge from inpatient rehabilitation. *J Head Trauma Rehabil* 1999; 14(3): 247-256.
- ²¹⁷ Gouvier W. D., Blanton P. D., LaPorte K. K., Nepomuceno C.: Reliability and validity of the disability rating scale and the levels of cognitive functioning scale in monitoring recovery from severe head injury. *Arch Phys Med Rehabil* 1987; 68: 94-97.
- ²¹⁸ Novack T. A., Bergquist T. F., Bennett G., Gouvier W. D.: Primary caregiver distress following severe head injury. *J Head Trauma Rehabil* 1992; 6(4): 69-77.
- ²¹⁹ Rappaport M., Hall K. M., Hopkins H. K. et al.: Disability rating scale for severe head trauma: coma to community. *Arch Phys Med Rehabil* 1982; 63: 118-123.
- ²²⁰ Hall K. M., Hamilton B. B., Gordon W. A., Zasler N.: Characteristics and comparisons of functional assessment indices: disability rating scale, functional independence measure, and functional assessment measure. *J Head Trauma Rehabil* 1993; 8: 60-74.
- ²²¹ Hall K., Cope D. N., Rappaport M.: Glasgow Outcome Scale and Disability Rating Scale: comparative usefulness in following recovery in traumatic head injury. *Arch Phys Med Rehabil* 1985; 66(1): 35-37.

-
- ²²² Ashley M. J., Persel C. S., Clark M. C., Krych D.K.: Long-term follow-up of post-acute traumatic brain injury rehabilitation: a statistical analysis to test for stability and predictability of outcome. *Brain Inj* 1997; 11(9): 677-690.
- ²²³ Eliason M. R., Topp B.W.: Predictive validity of Rappaport's Disability Rating Scale in subjects with acute brain dysfunction. *Phys Ther* 1984; 64(9): 1357-1360.
- ²²⁴ Boake C., High W. M.: Functional outcome from traumatic brain injury: unidimensional or multidimensional? *Am J Phys Med Rehabil* 1996; 75(2): 105-113.
- ²²⁵ Cifu D. X., Keyser-Marcus L., Lopez E., Wehman P. et al.: Acute predictors of successful return to work 1 year after traumatic brain injury: a multi-center analysis. *Arch Phys Med Rehabil* 1997; 78(2): 125-131.
- ²²⁶ Leung K. L., Man D. W.: Prediction of vocational outcome of people with brain injury after rehabilitation: a discriminant analysis. *Work* 2005; 25(4): 333-340.
- ²²⁷ Zhang L., Abreu B. C., Gonzales V., Seale G. et al.: Comparison of the Community Integration Questionnaire, the Craig Handicap Assessment and Reporting Technique, and the Disability Rating Scale in Traumatic Brain Injury. *J Head Trauma Rehabil* 2002; 17(6): 497-509.
- ²²⁸ Dikmen S., Machamer J., Miller B., Doctor J.: Functional Status Examination: A New Instrument for Assessing Outcome in Traumatic Brain Injury. *Journal of Neurotrauma* 2001; 18(2): 127-141.
- ²²⁹ Hudak A. M., Caesar R. R., Frol A. B., Krueger K. et al.: Functional outcome scales in traumatic brain injury: a comparison of the Glasgow Outcome Scale (Extended) and the Functional Status Examination. *Journal of Neurotrauma* 2005; 22(11): 1319-1326.
- ²³⁰ Ashley M. J., Persel C. S., Clark M. C.: Validation of an independent living scale for post-acute rehabilitation applications. *Brain Injury* 2001; 15(5): 435-442.
- ²³¹ Ashley M. J., Persel C. S., Clark M. C., Krych D. K. :Long-term follow-up of post-acute traumatic brain injury rehabilitation: a statistical analysis to test for stability and predictability of outcome. *Brain Injury* 1997; 11(9): 677-690.
- ²³² Dijkers M.: Measuring the long-term outcomes of traumatic brain injury: a review of Community Integration Questionnaire studies. *Journal of Head Trauma Rehabilitation* 1997; 12: 74-91.
- ²³³ Tepper S., Beatty P., DeJong G.: Outcomes in traumatic brain injury: self-report versus report of significant others. *Brain Injury* 1996; 10: 575-581.
- ²³⁴ Willer B., Rosenthal M., Kreutzer J. S., Gordon W. A. et al.: Assessment of community integration following rehabilitation for traumatic brain injury. *Journal of Head Trauma Rehabilitation* 1993; 8: 75-87.
- ²³⁵ Willer B., Linn R., Allen K. (1995): Community integration and barriers to integration for individuals with brain injury. In: Finlayson MAJ, Garner SH, editors. Brain injury rehabilitation: clinical considerations. Baltimore: Williams and Wilkins. 355-375.

-
- ²³⁶ Corrigan J. D., Deming R.: Psychometric characteristics of the Community Integration Questionnaire: Replication and extension. *Journal of Head Trauma Rehabilitation* 1995; 10: 41-53.
- ²³⁷ Sander A. M., Fuchs K. L., High W. M., Hall K. M. et al.: The Community Integration Questionnaire revisited: assessment of factor structure and validity. *Arch Phys Med Rehabil* 1999; 80: 1303-1308.
- ²³⁸ Hall K. M., Mann N., High W., Wright J. et al.: Functional measures after traumatic brain injury: ceiling effects of FIM, FIM+FAM, DRS and CIQ. *J Head Trauma Rehabil* 1996; 11(5): 27-39.
- ²³⁹ Willer B., Ottenbacher K. J., Coad M. L. The Community Integration Questionnaire: a comparative examination. *American Journal of Physical Medicine and Rehabilitation* 1994; 73: 103-111.
- ²⁴⁰ Willer B., Button J., Rempel R.: Residential and home-based postacute rehabilitation of individuals with traumatic brain injury: a case control study. *Archives of Physical Medicine and Rehabilitation* 1999; 80: 399-406.
- ²⁴¹ Kuipers P., Kendall M., Fleming J., Tate R.: Comparison of the Sydney Psychosocial Reintegration Scale (SPRS) with the Community Integration Questionnaire (CIQ): psychometric properties. *Brain Injury* 2004; 18(2): 161-177.
- ²⁴² Sherwin E., Whiteneck G., Corrigan J., Bedell G. et al.: Domains of a TBI minimal data set: community reintegration phase. *Brain Injury* 2006; 20(4): 383-389.
- ²⁴³ Johnston M. V., Shawaryn M. A., Malec J., Kreutzer J. et al.: The structure of functional and community outcomes following traumatic brain injury. *Brain Injury* 2006; 20(4): 391-407.
- ²⁴⁴ Wasser Th., Pasquale M. A., Matchett St., Bryan Y et al.: Establishing reliability and validity of the Critical Care Family Satisfaction Survey. *Crit Care Med* 2001; 29(1): 192-206.
- ²⁴⁵ Underhill A. T., LoBello S. G., Fine P. R.: Reliability and validity of the Family Satisfaction Scale with survivors of traumatic brain injury. *Journal of Rehabilitation Research & Development* 2004; 41(4): 603-610.
- ²⁴⁶ Lam T., Noonan V. K., Eng J. J.: A systematic review of functional ambulation outcome measures in spinal cord injury. *Spinal Cord* 2008; 46: 246-254.
- ²⁴⁷ Masur H. (2000): Skalen und Scores in der Neurologie, 2. Auflage. Georg Thieme Verlag. Stuttgart. S. 18